bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Decomposing variability in protein levels from noisy expression, genome duplication and partitioning errors during cell-divisions

M. Soltani[1], C. A. Vargas-Garcia[1], D. Antunes[2], A. Singh[1,3,4,5*]

1 Electrical and Computer Engineering, University of Delaware, Newark, DE.
2 Mechanical Engineering, Eindhoven University of Technology, Netherlands.
3 Biomedical Engineering, University of Delaware, Newark, DE.
4 Mathematical Sciences, University of Delaware, Newark, DE.
5 Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE.

∗ E-mail: Corresponding absingh@udel.udu

Abstract

Inside individual cells, expression of genes is inherently stochastic and manifests as cell-to-cell variability or noise in protein copy numbers. Since proteins half-lives can be comparable to the cell-cycle length, randomness in cell-division times generates additional intercellular variability in protein levels. Moreover, as many mRNA/protein species are expressed at low-copy numbers, errors incurred in partitioning of molecules between the mother and daughter cells are significant. We derive analytical formulas for the total noise in protein levels for a general class of cell-division time and partitioning error distributions. Using a novel hybrid approach the total noise is decomposed into components arising from i) stochastic expression; ii) partitioning errors at the time of cell-division and iii) random cell-division events. These formulas reveal that random cell-division times not only generate additional extrinsic noise but also critically affect the mean protein copy numbers and intrinsic noise components. Counter intuitively, in some parameter regimes noise in protein levels can decrease as cell-division times become more stochastic. Computations are extended to consider genome duplication, where the gene dosage is increased by two-fold at a random point in the cell-cycle. We systematically investigate how the timing of genome duplication influences different protein noise components. Intriguingly, results show that noise contribution from stochastic expression is minimized at an optimal genome duplication time. Our theoretical results motivate new experimental methods for decomposing protein noise levels from single-cell expression data. Characterizing the contributions of individual noise mechanisms will lead to precise estimates of gene expression parameters and techniques for altering stochasticity to change phenotype of individual cells.

Key index words or phrases

Single cell, stochastic gene expression, cell-division, moment dynamics, hybrid models, noise decomposition, partitioning errors

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Introduction

The level of a protein can deviate considerably from cell-to-cell, in spite of the fact that cells are genetically-identical and are in the same extracellular environment [1–3]. This intercellular variation or noise in protein counts has been implicated in diverse processes such as corrupting functioning of gene networks [4–6], driving probabilistic cell-fate decisions [7–12], buffering cell populations from hostile changes in the environment [13–16], and causing clonal cells to respond differently to the same stimulus [17–19]. An important source of noise driving random fluctuations in protein levels is stochastic gene expression due to the inherent probabilistic nature of biochemical processes [20–23]. Recent experimental studies have uncovered additional noise sources that affect protein copy numbers. For example, the time take to complete cell-cycle (i.e., time between two successive cell-division events) has been observed to be stochastic across organisms [24–32]. Given that many proteins/mRNAs are present inside cells at low-copy numbers, errors incurred in partitioning of molecules between the mother and daughter cells are significant [33–35]. Finally, the time at which a particular gene of interest is duplicated can also vary between cells [36, 37]. We investigate how such noise sources in the cell-cycle process combine with stochastic gene expression to generate intercellular variability in protein copy numbers (Fig.1).

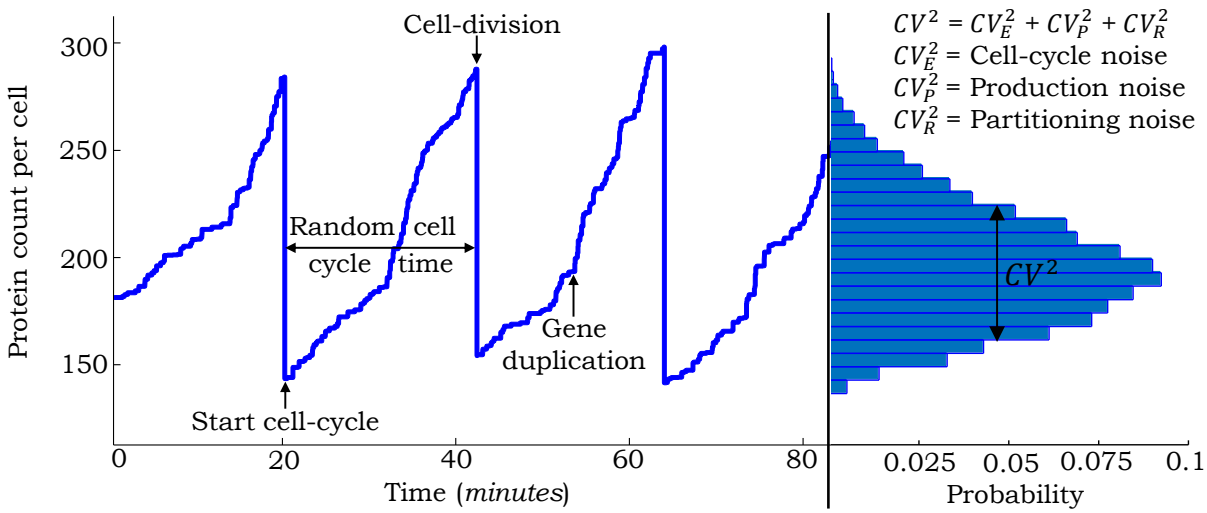

Figure 1. Sample trajectory of the protein level in a single cell with different sources of noise. Stochastically expressed proteins accumulate within the cell at a certain rate. At a random point in the cell-cycle, gene-duplication results in an increase in production rate. Stochastic cell-division events lead to random partitioning of protein molecules between the mother and daughter cells with each cell receiving, on average, half the number of proteins in the mother cell just before division. The steady-state protein copy number distribution obtained from a large number of trajectories is shown on the right. The total noise in the protein level, as measured by the Coefficient of Variation (CV) squared can be broken into contributions from individual noise mechanisms.

Prior studies that quantify the effects of cell-division on the protein noise level have been restricted to specific cases. For example, noise computations have been done in stochastic gene expression models, where cell-divisions occur at deterministic time intervals [33, 38, 39]. Recently, we have analyzed a deterministic model of gene expression with random cell-division events [40]. Building up on this work, we formulate a mathematical model that couples stochastic expression of a stable protein with random cell-division events that follow an arbitrary probability distribution function. Moreover, at the time of cell-division, proteins are randomly partitioned between the mother and daughter cells based on a gen-

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

eral framework that allows the partitioning errors to be higher or lower than as predicted by binomial partitioning. For this class of models, we derive an exact analytical formula for the protein noise level as quantified by the steady-state Coefficient of Variation (CV) squared. This formula is further decomposed into individual components representing contributions from different noise sources. A systematic investigation of this formula leads to novel insights, such as identification of regimes where increasing randomness in the timing of cell-division events decreases the protein noise level.

Next, we extend the above model to include genome duplication events that increase the gene's transcription rate by two-fold (corresponding to doubling of gene dosage) prior to cell-division [36, 41]. To our knowledge, this is the first study integrating randomness in the genome duplication process with stochastic gene expression. An exact formula for the protein noise level is derived for this extended model and used to investigate how the timing of duplication affects different noise components. Counter intuitively, results show that doubling of the transcription rate within the cell-cycle can lead to smaller fluctuations in protein levels as compared to a constant transcription rate through out the cell-cycle. Finally, we discuss how formulas obtained in this study can be used to infer parameters and characterize the gene expression process from single-cell studies.

2 Coupling gene expression to cell-division

We consider the standard model of stochastic gene expression [42, 43], where mRNAs are transcribed at exponentially distributed time intervals from a constitutive gene with rate k_x. For the time being, we exclude genome duplication and the transcription rate is fixed throughout the cell-cycle. Assuming short-lived mRNAs, each transcription event results in a burst of proteins [43–45]. The corresponding jump in protein levels is shown as

$$x(t) \mapsto x(t) + B, \tag{1}$$

where $x(t)$ is the protein population count in the mother cell at time t, B is a random burst size drawn from a positively-valued distribution and represents the number of protein molecules synthesized in a single-mRNA lifetime. Motivated by observations in *E. coli* and mammalian cells, where many proteins have half-lives considerably longer than the cell-doubling time, we assume a stable protein with no active degradation [46–48]. Thus, proteins accumulate within the cell till the time of cell-division, at which point they are randomly partitioned between the mother and daughter cells.

Let cell-division events occur at times $t_s,\ s \in \{1, 2, \ldots\}$. The cell-cycle time

$$T := t_s - t_{s-1}, \tag{2}$$

follows an arbitrary positively-valued probability distribution with the following mean and coefficient of variation (CV) squared

$$\langle T \rangle = \langle t_s - t_{s-1} \rangle, \quad CV_T^2 = \frac{\langle T^2 \rangle - \langle T \rangle^2}{\langle T \rangle^2}, \tag{3}$$

where $\langle . \rangle$ denotes expected value through out this paper. The random change in $x(t)$ during cell-division is given by

$$x(t_s) \mapsto x_+(t_s), \tag{4}$$

where $x(t_s)$ and $x_+(t_s)$ denote the protein levels in the mother cell just before and after division, respectively. Conditioned on $x(t_s)$, $x_+(t_s)$ is assumed to have the following statistics

$$\langle x_+(t_s) | x(t_s) \rangle = \frac{x(t_s)}{2}, \quad \left\langle x_+^2(t_s) - \langle x_+(t_s) \rangle^2 \middle| x(t_s) \right\rangle = \frac{\alpha x(t_s)}{4}. \tag{5}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The first equation implies symmetric division, i.e., on average the mother cell inherits half the number protein molecules just before division. The second equation in (5) describes the variance of $\langle x_+(t_s)\rangle$ and quantifies the error in partitioning of molecules through the non-negative parameter α. For example, $\alpha = 0$ represents deterministic partitioning where $x_+(t_s) = x(t_s)/2$ with probability equal to one. A more realistic model for partitioning is each molecule having an equal probability of being in the mother or daughter cell [49–51]. This result in a binomial distribution for $x_+(t_s)$

$$\text{Probability}\{x_+(t_s) = j|x(t_s)\} = \frac{x(t_s)!}{j!(x(t_s)-j)!}\left(\frac{1}{2}\right)^{x(t_s)}, \quad j \in \{0, 1, \ldots, x(t_s)\}, \tag{6}$$

and corresponds to $\alpha = 1$ in (5). Interestingly, recent studies have shown that partitioning of proteins that form clusters or multimers can result in $\alpha > 1$ in (5), i.e., partitioning errors are much higher than as predicted by the binomial distribution [33, 39]. In contrast, if molecules push each other to opposite poles of the cell, then the partitioning errors will be smaller than as predicted by (6) and $\alpha < 1$.

The model with all the different noise mechanisms (stochastic expression; random cell-division events and partitioning errors) is illustrated in Fig. 2A and referred to as the full model. We also introduce two additional hybrid models [52, 53], where protein production and partitioning are considered in their deterministic limit (Fig. 2B-C). Note that unlike the full model, where $x(t)$ takes non-negative integer values, $x(t)$ is continuous in the hybrid models. We will use these hybrid models for decomposing the protein noise level obtained from the full model into individual components representing contributions from different noise sources. However, before computing the noise, we first determine the average number of proteins as a function of the cell-cycle time distribution.

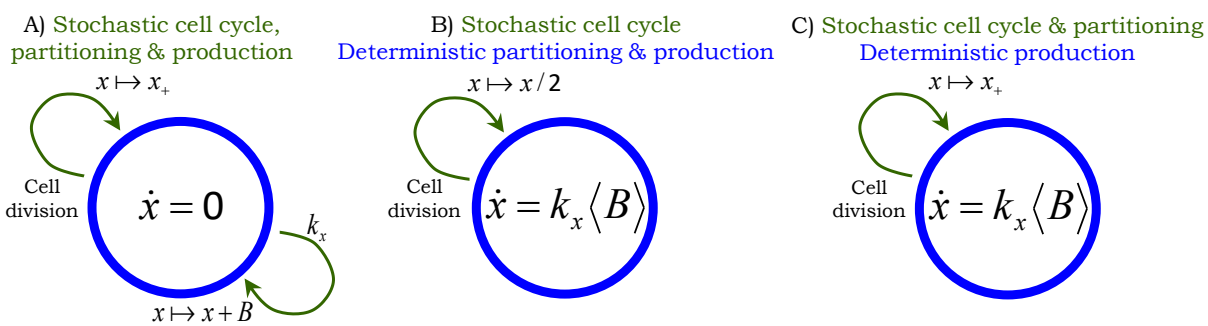

Figure 2. Stochastic models of gene expression with cell-division. Arrows denote stochastic events that change the protein level by discrete jumps as shown in (1) and (4). The differential equation within the circle represents the time evolution of $x(t)$ in between events. **A)** Model with all the different sources of noise: proteins are expressed in stochastic bursts, cell-division occurs at random times, and molecules are partitioned between the mother and daughter cells based on (5). The trivial dynamics $\dot{x} = 0$ signifies that the protein level is constant in-between stochastic events. **B)** Hybrid model where randomness in cell-division events is the only source of noise. Protein production is modeled deterministically through a differential equation and partitioning errors are absent, i.e., $\alpha = 0$ in (5). **C)** Hybrid model where noise comes from both cell-division events and partitioning errors. Protein production is considered deterministically as in Fig. 2B. Since $x(t)$ is continuous here, $x_+(t_s)$ has a positively-valued continuous distribution with same mean and variance as in (5)

3 Computing the average number of protein molecules

To quantify the steady-state mean protein level we consider the full model illustrated in Fig. 2A. It turns out that all the models shown in Fig. 2 are identical in terms of finding $\langle x(t)\rangle$ and in principle

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

any one of them could have been used. To obtain differential equations describing the time evolution of $\langle x(t)\rangle$ we model the cell-cycle time through a phase-type distribution, which can be represented by a continuous-time Markov chain. Phase-type distributions are dense in the class of positively-valued continuous distributions, i.e., one can always construct a sequence of phase-type distributions that converges point wise to a given distribution of interest [54]. We use this denseness property as a practical tool for modeling the cell-cycle time.

3.1 Cell-cycle time as a phase-type distribution

We consider a class of phase-type distribution that consists of a mixture of Erlang distributions. Recall that an Erlang distribution of order i is the distribution of the sum of i independent and identical exponential random variables. The cell-cycle time is assumed to have an Erlang distribution of order i with probability $p_i,\ i = \{1,\dots,n\}$ and can be represented by a continuous-time Markov chain with states $G_{ij},\ j = \{1,\dots,i\},\ i = \{1,\dots,n\}$ (Fig. 3). Let Bernoulli random variables $g_{ij} = 1$ if the system resides in state G_{ij} and 0 otherwise. The probability of transition $G_{ij} \to G_{i(j+1)}$ in the next infinitesimal time interval $[t, t+dt)$ is given by $kg_{ij}dt$, implying that the time spent in each state G_{ij} is exponentially distributed with mean $1/k$. To summarize, at the start of cell-cycle, a state $G_{i1},\ i = \{1,\dots,n\}$ is chosen with probability p_i and cell-division occurs after transitioning through i exponentially distributed steps. Based on this formulation, the probability of a cell-division event occurring in the next time interval $[t, t+dt)$ is given by $kp_i \sum_{j=1}^{n} g_{jj}dt$, and whenever the event occurs, the protein level changes as per (4). Finally, the mean and the coefficient of variation squared of the cell-cycle time is obtained as

$$\langle T\rangle = \sum_{i=1}^{n} p_i \frac{i}{k}, \quad CV_T^2 = \frac{1}{k}\frac{1}{\langle T\rangle} \tag{7}$$

in terms of the Markov chain parameters. Our goal is to obtain $\overline{\langle x\rangle} := \lim_{t\to\infty}\langle x(t)\rangle$ as a function of $\langle T\rangle$ and CV_T^2.

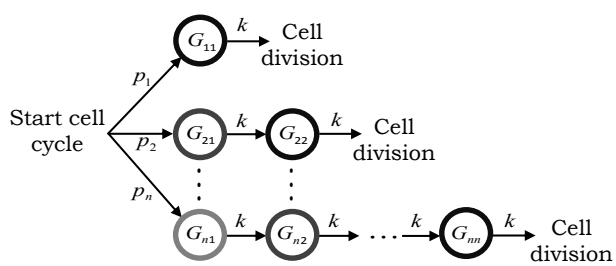

Figure 3. A continuous-time Markov chain model for the cell-cycle time. The cell-cycle time is assumed to follow a mixture of Erlang distributions. At the start of cell-cycle, a state $G_{i1},\ i = \{1,\dots,n\}$ is chosen with probability p_i. The cell-cycle transitions through states $G_{ij},\ j = \{1,\dots,i\}$ residing for an exponentially distributed time with mean $1/k$ in each state. Cell-division occurs after exit from G_{ii} and the above process is repeated.

3.2 Time evolution of the mean protein level

Time evolution of the statistical moments of $x(t)$ can be obtained from the Kolmogorov forward equations corresponding to the full model in Fig. 2A combined with the cell-division process described in Fig. 3. We refer the reader to [52, 55, 56] for an introduction to moment dynamics for stochastic and hybrid systems. Analysis in Appendix A shows

$$\frac{d\langle x\rangle}{dt} = k_x\langle B\rangle - \frac{k}{2}\left\langle \sum_{j=1}^{n} x g_{jj} \right\rangle. \tag{8}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Note that the time-derivative of the mean protein level (first-order moment) is unclosed, in the sense that, it depends on the second-order moment $\langle x g_{ij} \rangle$. Typically, approximate closure methods are used to solve moments in such cases [52,56–61]. However, the fact that g_{ij} is binary can be exploited to automatically close moment dynamics. In particular, since $g_{ij} \in \{0, 1\}$

$$\langle g_{ij}^n x^m \rangle = \langle g_{ij} x^m \rangle, \quad n \in \{1, 2, \ldots\} \tag{9}$$

for any non-negative integer m. Moreover, as only a single state g_{ij} can be 1 at any time

$$\langle g_{ij} g_{rq} x^m \rangle = 0, \text{ if } i \neq r \text{ or } j \neq q. \tag{10}$$

Using (9) and (10), the time evolution of $\langle x g_{ij} \rangle$ is obtained as

$$\frac{d\langle x g_{i1} \rangle}{dt} = \frac{k_x \langle B \rangle p_i}{i} + \frac{k}{2} p_i \left\langle \sum_{j=1}^{n} x g_{jj} \right\rangle - k \langle x g_{i1} \rangle, \tag{11a}$$

$$\frac{d\langle x g_{ij} \rangle}{dt} = \frac{k_x \langle B \rangle p_i}{i} - k \langle x g_{ij} \rangle + k \langle x g_{i(j-1)} \rangle, \; j = \{2, \ldots, i\} \tag{11b}$$

and only depends on $\langle x g_{ij} \rangle$ (see Appendix A). Thus, (8) and (11) constitute a closed system of linear differential equations from which moments can be computed exactly.

To obtain an analytical formula for the average number of proteins, we start by performing a steady-state analysis of (8) that yields

$$\overline{\left\langle \sum_{j=1}^{n} x g_{jj} \right\rangle} = \frac{2 k_x \langle B \rangle}{k}, \tag{12}$$

where $\overline{\langle . \rangle}$ denotes the expected value in the limit $t \to \infty$. Using (12), $\overline{\langle x g_{i1} \rangle}$ is determined from (11a), and then all moments $\overline{\langle x g_{ij} \rangle}$ are obtained recursively by performing a steady-state analysis of (11b) for $j = \{2, \ldots, i\}$. This analysis results in

$$\overline{\langle x g_{ij} \rangle} = \frac{k_x \langle B \rangle}{k} p_i \left(1 + \frac{j}{i} \right). \tag{13}$$

Using (7), (13) and the fact that $\sum_{i=1}^{n} \sum_{j=1}^{i} g_{ij} = 1$ we obtain the following expression for the mean protein level

$$\overline{\langle x \rangle} = \overline{\left\langle x \sum_{i=1}^{n} \sum_{j=1}^{i} g_{ij} \right\rangle} = \sum_{i=1}^{n} \sum_{j=1}^{i} \overline{\langle x g_{ij} \rangle} = \frac{k_x \langle B \rangle \langle T \rangle \left(3 + CV_T^2 \right)}{2}. \tag{14}$$

It is important to point that (14) holds irrespective of the complexity, i.e., the number of sates G_{ij} used in the phase-type distribution to approximate the cell-cycle time distribution. As expected, $\overline{\langle x \rangle}$ increases linearly with the average cell-cycle time duration $\langle T \rangle$ with longer cell-cycles resulting in more accumulation of proteins. Consistent with previous findings, (14) shows that the mean protein level is also affected by the randomness in the cell-cycle times (CV_T^2) [40,62]. For example, $\overline{\langle x \rangle}$ reduces by 25% as T changes from being exponentially distributed ($CV_T^2 = 1$) to periodic ($CV_T^2 = 0$) for fixed $\langle T \rangle$ fixed. Next, we determine the noise in protein copy numbers, as quantified by the coefficient of variation squared.

4 Computing the protein noise level

Recall that the full model introduced in Fig. 2A has three distinct noise mechanisms. Our strategy for computing the protein noise level is to first analyze the model with a single noise source, and then consider models with two and three sources. As shown below, this approach provides a systematic dissection of the protein noise level into components representing contributions from different mechanisms.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.1 Contribution from randomness in cell-cycle times

We begin with the model shown in Fig. 2B, where noise comes from a single source - random cell-division events. For this model, the time evolution of the second-order moment of the protein copy number is obtained as

$$\frac{d\langle x^2\rangle}{dt} = 2k_x\langle B\rangle\langle x\rangle - \frac{3k}{4}\left\langle \sum_{j=1}^{n} x^2 g_{jj}\right\rangle, \tag{15}$$

and depends on third-order moments $\langle x^2 g_{jj}\rangle$ (see Appendix B). Using the approach introduced earlier for obtaining the mean protein level, we close moment equations by writing the time evolution of moments $\langle x^2 g_{ij}\rangle$. Using (9) and (10)

$$\frac{d\langle x^2 g_{i1}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{i1}\rangle + \frac{k}{4}p_i\left\langle \sum_{j=1}^{n} x^2 g_{jj}\right\rangle - k\langle x^2 g_{i1}\rangle, \tag{16a}$$

$$\frac{d\langle x^2 g_{ij}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{ij}\rangle - k\langle x^2 g_{ij}\rangle + k\langle x^2 g_{(i-1)j}\rangle,\ j = \{2,\ldots,i\}\,. \tag{16b}$$

Note that the moment dynamics for $\langle x\rangle$ and $\langle x g_{ij}\rangle$ obtained in the previous section (equations (8) and (11)) are identical for all the models in Fig. 2, irrespective of whether the noise mechanism is modeled deterministically or stochastically. Equations (8), (11), (15) and (16) represent a closed set of linear differential equations and their steady-state analysis yields

$$\overline{\langle x^2 g_{ij}\rangle} = \frac{k_x^2\langle B\rangle^2\langle T\rangle\left(3+CV_T^2\right)}{3k}p_i + \frac{2k_x^2\langle B\rangle^2}{k}\left(\frac{j^2+j}{i}\right)p_i. \tag{17}$$

From (96)

$$\overline{\langle x^2\rangle} = \overline{\left\langle x^2\sum_{i=1}^{n}\sum_{j=1}^{i} g_{ij}\right\rangle} = \sum_{i=1}^{n}\sum_{j=1}^{i}\overline{\langle x^2 g_{ij}\rangle} = k_x^2\langle B\rangle^2\frac{\langle T^3\rangle + 4CV_T^2\langle T\rangle^3 + 6\langle T\rangle^3}{3\langle T\rangle}, \tag{18a}$$

$$\langle T^3\rangle = \frac{2}{k}CV_T^2 + \frac{3}{k}\langle T\rangle^2 + \langle T\rangle^3, \tag{18b}$$

where $\langle T^3\rangle$ is the third-order moment of the cell-cycle time. Using (18) and the mean protein count quantified in (14), we obtain the following coefficient of variation squared

$$CV_E^2 = \frac{1}{27} + \frac{4\left(9\frac{\langle T^3\rangle}{\langle T\rangle^3} - 9 - 6CV_T^2 - 7CV_T^4\right)}{27\left(3+CV_T^2\right)^2}, \tag{19}$$

which represents the noise contribution from random cell-division events. Since cell-division is a global event that affects expression of all genes, this noise contribution can also be referred to as *extrinsic noise* [49, 63–66]. In reality, there would be other sources of extrinsic noise, such as, fluctuations in the gene-expression machinery that we have ignored in this analysis.

Note that $CV_E^2 \to 1/27$ as T approaches a delta distribution, i.e., cell divisions occur at fixed time intervals. We discuss simplifications of (19) in various limits. For example, if the time taken to complete cell-cycle is lognormally distributed, then

$$\frac{\langle T^3\rangle}{\langle T\rangle^3} = \left(1+CV_T^2\right)^3 \implies CV_E^2 = \frac{1}{27} + \frac{4\left(21CV_T^2 + 20CV_T^4 + 9CV_T^6\right)}{27\left(3+CV_T^2\right)^2} \tag{20}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and extrinsic noise monotonically increases with CV_T^2. If fluctuations in T around $\langle T\rangle$ are small, then using Taylor series

$$\langle T^3\rangle/\langle T\rangle^3 \approx 1 + 3CV_T^2. \tag{21}$$

Substituting (21) in (19) and ignoring CV_T^4 and higher order terms yields

$$CV_E^2 \approx \frac{1}{27} + \frac{28CV_T^2}{81}, \tag{22}$$

where the first term is the extrinsic noise for $CV_T^2 \to 0$ and the second term is the additional noise due to random cell-division events.

4.2 Contribution from partitioning errors

Next, we consider the model illustrated in Fig. 2C with both random cell-division events and partitioning of protein between the mother and daughter cells. Thus, the protein noise level here represents the contribution from both these sources. Analysis in Appendix C shows that the time evolution of $\langle x^2\rangle$ and $\langle x^2 g_{ij}\rangle$ are given by

$$\frac{d\langle x^2\rangle}{dt} = 2k_x\langle B\rangle\langle x\rangle + \frac{1}{4}\alpha k\left\langle \sum_{j=1}^{n} x g_{jj}\right\rangle - \frac{3}{4}k\left\langle \sum_{j=1}^{n} x^2 g_{jj}\right\rangle, \tag{23a}$$

$$\frac{d\langle x^2 g_{i1}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{i1}\rangle + \frac{k}{4}p_i\left\langle \sum_{j=1}^{n} x^2 g_{jj}\right\rangle + \frac{1}{4}\alpha k p_i\left\langle \sum_{j=1}^{n} x g_{jj}\right\rangle - k\langle x^2 g_{i1}\rangle, \tag{23b}$$

$$\frac{d\langle x^2 g_{ij}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{ij}\rangle - k\langle x^2 g_{ij}\rangle + k\langle x^2 g_{(i-1)j}\rangle,\ j = \{2,\ldots,i\}. \tag{23c}$$

Note that (23a)-(23b) are slightly different from their counterparts obtained in the previous section (equations (15) and (16a)) with additional terms that depend on α, where α quantifies the degree of partitioning error as defined in (5). As expected, (23) reduces to (15)-(16) when $\alpha = 0$ (i.e., deterministic partitioning). Computing $\overline{\langle x^2 g_{ij}\rangle}$ by performing a steady-state analysis of (23) and using a similar approach as in (18) we obtain

$$\overline{\langle x^2\rangle} = \frac{\langle T^3\rangle + 4CV_T^2\langle T\rangle^3 + 6\langle T\rangle^3}{3\langle T\rangle} + \frac{2\alpha k_x\langle B\rangle\langle T\rangle}{3}. \tag{24}$$

Finding CV^2 of the protein level and subtracting the extrinsic noise found in (19) yields

$$CV_R^2 = \frac{4\alpha}{3(3+CV_T^2)}\frac{1}{\overline{\langle x\rangle}}, \tag{25}$$

where CV_R^2 represents the contribution of partitioning errors to the protein noise level. Intriguingly, while CV_R^2 increases with α, it decrease with CV_T^2. Thus, as cell-division times become more random for a fixed $\langle T\rangle$ and $\overline{\langle x\rangle}$, the noise contribution from partitioning errors decrease.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4.3 Contribution from stochastic expression

Finally, we consider the full model in Fig. 2A with all the three different noise sources. For this model, moment dynamics is obtained as (see Appendix D)

$$\frac{d\langle x^2\rangle}{dt} = k_x\langle B^2\rangle + 2k_x\langle B\rangle\langle x\rangle + \frac{1}{4}\alpha k\left\langle \sum_{j=1}^{n} x g_{jj}\right\rangle - \frac{3k}{4}\left\langle \sum_{j=1}^{n} x^2 g_{jj}\right\rangle, \tag{26a}$$

$$\frac{d\langle x^2 g_{i1}\rangle}{dt} = \frac{k_x\langle B^2\rangle p_i}{i} + 2k_x\langle B\rangle\langle x g_{i1}\rangle + \frac{k}{4}p_i\left\langle \sum_{j=1}^{n} x^2 g_{jj}\right\rangle + \frac{1}{4}\alpha k p_i\left\langle \sum_{j=1}^{n} x g_{jj}\right\rangle - k\langle x^2 g_{i1}\rangle, \tag{26b}$$

$$\frac{d\langle x^2 g_{ij}\rangle}{dt} = \frac{k_x\langle B^2\rangle p_i}{i} + 2k_x\langle B\rangle\langle x g_{ij}\rangle - k\langle x^2 g_{ij}\rangle + k\langle x^2 g_{(i-1)j}\rangle,\ j = \{2,\ldots,i\}\,. \tag{26c}$$

Compared to (23), (26) has additional terms of the form $k_x\langle B^2\rangle$, where $\langle B^2\rangle$ is the second-order moment of the protein burst size in (1). Performing an identical analysis as before we obtain

$$\overline{\langle x^2\rangle} = \frac{\langle T^3\rangle + 4CV_T^2\langle T\rangle^3 + 6\langle T\rangle^3}{3\langle T\rangle} + \frac{2\alpha k_x\langle B\rangle\langle T\rangle}{3} + \frac{k_x\langle B^2\rangle\langle T\rangle(3CV_T^2+5)}{2}, \tag{27}$$

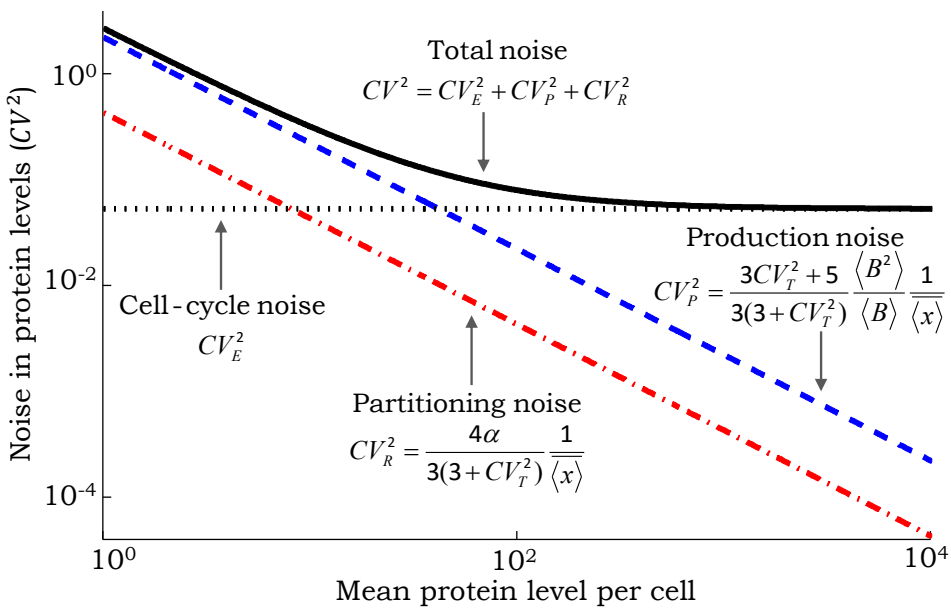

Figure 4. Scaling of noise as a function of the mean protein level for different mechanisms. The contribution of random cell-division events to the noise in protein copy numbers (extrinsic noise) is invariant of the mean. In contrast, contributions from partitioning errors at the time of cell-division (partitioning noise) and stochastic expression (production noise) scale inversely with the mean. The scaling factors are shown as a function of the protein random burst size B, noise in cell-cycle time (CV_T^2) and magnitude of partitioning errors quantified by α (see (5)). With increasing mean level the total noise first decreases and then reaches a baseline that corresponds to extrinsic noise. For this plot, B is assumed to be geometrically-distributed with mean $\langle B\rangle = 1.5$, $CV_T^2 = 0$ and $\alpha = 1$ (i.e., binomial partitioning).

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

which yields the following total protein noise level

$$CV^2 = CV_E^2 + CV_R^2 + CV_P^2 = CV_E^2 + \underbrace{\overbrace{\frac{4\alpha}{3(3+CV_T^2)}\frac{1}{\overline{\langle x\rangle}}}^{\text{Partitioning noise } (CV_R^2)} + \overbrace{\frac{3CV_T^2+5}{3(3+CV_T^2)}\frac{\langle B^2\rangle}{\langle B\rangle}\frac{1}{\overline{\langle x\rangle}}}^{\text{Production noise } (CV_P^2)}}_{\text{Intrinsic noise}}, \tag{28}$$

that can be decomposed into three terms. The first is the extrinsic noise CV_E^2 representing the contribution from random cell-division events and given by (19). The second term CV_R^2 is the contribution from partitioning errors determined in the previous section (partitioning noise), and the final term CV_P^2 is the additional noise representing the contribution from stochastic expression (production noise). We refer to the sum of the contributions from partitioning errors and stochastic expression as *intrinsic noise*. These intrinsic and extrinsic noise components are generally obtained experimentally using the dual-color assay that measures the correlation in the expression of two identical copies of the gene [49].

Interestingly, for a fixed mean protein level $\overline{\langle x\rangle}$, CV_T^2 has opposite effects on CV_R^2 and CV_P^2. While CV_R^2 monotonically decreases with increasing CV_T^2, CV_P^2 increases with CV_T^2. It turns out that in certain cases these effects can cancel each other out. For example, when $B = 1$ with probability one, i.e., proteins are synthesized one at a time at exponentially distributed time intervals and $\alpha = 1$ (binomial partitioning)

$$CV^2 = CV_E^2 + \frac{4}{3(3+CV_T^2)}\frac{1}{\overline{\langle x\rangle}} + \frac{3CV_T^2+5}{3(3+CV_T^2)}\frac{1}{\overline{\langle x\rangle}} = CV_E^2 + \frac{1}{\overline{\langle x\rangle}}. \tag{29}$$

In this limit the intrinsic noise is always 1/Mean irrespective of the cell-cycle time distribution T [33]. Note that the average number of proteins itself depends on T as shown in (14). Another important limit is $CV_T^2 \to 0$, in which case (28) reduces to

$$CV^2 \approx \underbrace{\overbrace{\frac{1}{27}}^{CV_E^2}}_{\text{Extrinsic noise}} + \underbrace{\overbrace{\frac{4\alpha}{9}\frac{1}{\overline{\langle x\rangle}}}^{CV_R^2} + \overbrace{\frac{5}{9}\frac{\langle B^2\rangle}{\langle B\rangle}\frac{1}{\overline{\langle x\rangle}}}^{CV_P^2}}_{\text{Intrinsic noise}}, \tag{30}$$

and is similar to the result obtained in [38] for deterministic cell-division times and binomial partitioning.

Fig. 4 shows how different protein noise components change as a function of the mean protein level as the gene's transcription rate k_x is modulated. The extrinsic noise is primarily determined by the distribution of the cell-cycle time and is completely independent of the mean. In contrast, both CV_R^2 and CV_P^2 scale inversely with the mean, albeit with different scaling factors (Fig. 4). This observation is particularly important since many single-cell studies in *E. coli*, yeast and mammalian cells have found the protein noise levels to scale inversely with the mean across different genes [67–70]. Based on this scaling it is often assumed that the observed cell-to-cell variability in protein copy numbers is a result of stochastic expression. However, as our results show, noise generated thorough partitioning errors is also consistent with these experimental observations and it may be impossible to distinguish between these two noise mechanisms based on protein CV^2 versus mean plots unless α is known.

5 Quantifying the effects of gene-duplication on protein noise

The full model introduced in Fig. 2 assumes that the transcription rate (i.e., the protein burst arrival rate) is constant throughout the cell-cycle. This model is now extended to incorporate gene duplication during cell cycle, which is assumed to create a two-fold change in the burst arrival rate (Fig. 5). As a result of this, accumulation of proteins will be bilinear as illustrated in Fig. 1. We divide the cell-cycle time T into two intervals: time from the start of cell-cycle to gene-duplication (T_1), and time from

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

gene-duplication to cell-division (T_2). T_1 and T_2 are independent random variables that follow arbitrary distributions modeled through phase-type processes (see Fig. S2 in the Supplementary Information). The mean cell-cycle duration and its noise can be expressed as

$$\langle T\rangle = \langle T_1\rangle + \langle T_2\rangle, \quad \beta = \frac{\langle T_1\rangle}{\langle T\rangle}, \quad CV_T^2 = \beta^2 CV_{T_1}^2 + (1-\beta)^2 CV_{T_2}^2, \tag{31}$$

where CV_X^2 denotes the coefficient of variation squared of the random variable X. An important variable in this formulation is β, which represents the average time of gene-duplication normalized by the mean cell-cycle time. Thus, β values close to 0 (1) imply that the gene is duplicated early (late) in the cell-cycle process. Moreover, the noise in the gene-duplication time is controlled via $CV_{T_1}^2$.

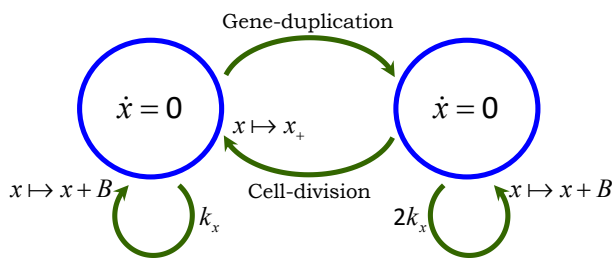

Figure 5. Model illustrating stochastic expression together with random gene-duplication and cell-division events. At the start of cell-cycle, protein production occurs in stochastic bursts with rate k_x. Genome duplication occurs at a random point T_1 within the cell-cycle and increases the burst arrival rate to $2k_x$. Cell-division occurs after time T_2 from genome duplication, at which point the burst arrival rate reverts back to k_x and proteins are randomly partitioned between cells based on (4).

We refer the reader to Appendix E for a detailed analysis of the model in Fig. 5 and only present the main results on the protein mean and noise levels. The steady-state mean protein count is given by

$$\overline{\langle x\rangle} = \frac{k_x\langle B\rangle\langle T_1\rangle\left(4-\beta+\beta CV_{T_1}^2\right)}{2} + k_x\langle B\rangle\langle T_2\rangle\left(3-\beta+(1-\beta)CV_{T_2}^2\right), \tag{32}$$

and decreases with β, i.e., a gene that duplicates early has on average, more number of proteins. When $\beta = 1$, then the transcription rate is k_x throughout the cell-cycle and we recover the mean protein level obtained in (14). Similarly, when $\beta = 0$ the transcription rate is $2k_x$ and we obtain twice the amount as in (14). As per our earlier observation, more randomness in the timing of genome duplication and cell-division (i.e., higher $CV_{T_1}^2$ and $CV_{T_2}^2$ values) increases $\overline{\langle x\rangle}$.

Our analysis shows that the total protein noise level can be decomposed into three components

$$CV^2 = CV_E^2 + CV_R^2 + CV_P^2 \tag{33}$$

where CV_E^2 is the extrinsic noise from random genome-duplication and cell-division events. Given its complexity, we refer the reader to equation (100) in Appendix E2 for an exact formula for CV_E^2. Moreover, the intrinsic noise, which represents the sum of contributions from partitioning errors (CV_R^2) and stochastic expression (CV_P^2) is obtained as

$$CV_R^2 + CV_P^2 = \overbrace{\frac{4\alpha(2-\beta)}{3\left((\beta^2-4\beta+6)+\beta^2 CV_{T_1}^2+2(1-\beta)^2 CV_{T_2}^2\right)}\frac{1}{\overline{\langle x\rangle}}}^{CV_R^2} + \overbrace{\frac{(10-8\beta+3\beta^2)+6(1-\beta)^2 CV_{T_2}^2+3\beta^2 CV_{T_1}^2}{3\left((\beta^2-4\beta+6)+\beta^2 CV_{T_1}^2+2(1-\beta)^2 CV_{T_2}^2\right)}\frac{\langle B^2\rangle}{\langle B\rangle}\frac{1}{\overline{\langle x\rangle}}}^{CV_P^2}. \tag{34}$$

Note that for $\beta = 0$ and 1, we recover the intrinsic noise level in (28) from (34). Interestingly, for $B = 1$ with probability 1 and $\alpha = 1$, the intrinsic noise is always 1/Mean irrespective of the values chosen for

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$CV^2_{T_1}$, $CV^2_{T_2}$ and β. For high precision in the timing of cell-cycle events ($CV_{T_1} \to 0, CV_{T_2} \to 0$)

$$CV^2 \approx \underbrace{\overbrace{\frac{4-3(\beta-2)^2\beta^2}{3\left(\beta^2-4\beta+6\right)^2}}^{CV^2_E}}_{\text{Extrinsic noise}} + \underbrace{\overbrace{\frac{4\alpha(2-\beta)}{3\left(\beta^2-4\beta+6\right)}\frac{1}{\overline{\langle x\rangle}}}^{CV^2_R} + \overbrace{\frac{(10-8\beta+3\beta^2)}{3\left(\beta^2-4\beta+6\right)}\frac{\langle B^2\rangle}{\langle B\rangle}\frac{1}{\overline{\langle x\rangle}}}^{CV^2_P}}_{\text{Intrinsic noise}}, \tag{35}$$

where mean protein level is given by

$$\overline{\langle x\rangle} \approx \frac{k_x\langle B\rangle\langle T_1\rangle\left(4-\beta\right)}{2} + k_x\langle B\rangle\langle T_2\rangle\left(3-\beta\right). \tag{36}$$

We investigate how different noise components in (35) vary with β as the mean protein level is held fixed by changing k_x. Fig. 6 shows that CV^2_P follows a U-shaped profile with the optima occurring at $\beta = 2-\sqrt{2} \approx 0.6$ and the corresponding minimum value being $\approx 5\%$ lower that its value at $\beta = 0$. An implication of this result is that if stochastic expression is the dominant noise source, then gene-duplication can result in slightly lower protein noise levels. In contrast to CV^2_P, CV^2_R has a maxima at $\beta = 2-\sqrt{2}$ which is $\approx 6\%$ higher than its value at $\beta = 0$ (Fig. 6). Analysis in Appendix E5 reveals that CV^2_R and CV^2_R follow the same qualitative shapes as in Fig. 6 for non-zero $CV^2_{T_1}$ and $CV^2_{T_2}$. Interestingly, when $CV^2_{T_1} = CV^2_{T_2}$, the maximum and minimum values of CV^2_R and CV^2_P always occur at $\beta = 2-\sqrt{2}$ albeit with different optimal values than Fig. 6 (see Fig. S3 in the Supplementary Information). For example, if $CV^2_{T_1} = CV^2_{T_2} = 1$ (i.e., exponentially distributed T_1 and T_2), then the maximum value of CV^2_R is 20% higher and the minimum value of CV^2_P is 10% lower than their respective value for $\beta = 0$. Given that the effect of changing β on CV^2_P and CV^2_R is small and antagonistic, the overall affect of genome duplication on intrinsic noise may be minimal and hard to detect experimentally.

6 Discussion

We have investigated a model of protein expression in bursts coupled to discrete gene-duplication and cell-division events. The novelty of our modeling framework lies in describing the size of protein bursts, T_1 (time between cell birth and gene duplication), T_2 (time between gene duplication and cell division) and partitioning of molecules during cell division through arbitrary distributions. Exact formulas connecting the protein mean and noise levels to these underlying distributions were derived. Furthermore, the protein noise level, as measured by the coefficient of variation squared, was decomposed into three components representing contributions from gene-duplication/cell-division events, stochastic expression and random partitioning. While the first component is independent of the mean protein level, the other two components are inversely proportional to it. Key insights obtained are as follows:

- The mean protein level is affected by both the first and second-order moments of T_1 and T_2. In particular, randomness in these times (for a fixed mean) increases the average protein count.
- Random gene-duplication/cell-division events create an extrinsic noise term which is completely determined by moments of T_1 and T_2 up to order three.
- The noise contribution from partitioning errors decreases with increasing randomness in T_1 and T_2. Thus, if $\overline{\langle x\rangle}$ is sufficiently small and α is large compared to B in (34), increasing noise in the timing of cell-cycle events decreases the total noise level.
- Genome duplication has counter intuitive effects on the protein noise level (Fig. 6). For example, if stochastic expression is the dominant source of noise, then doubling of transcription due to duplication results in lower noise as compared to constant transcription throughout the cell-cycle.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

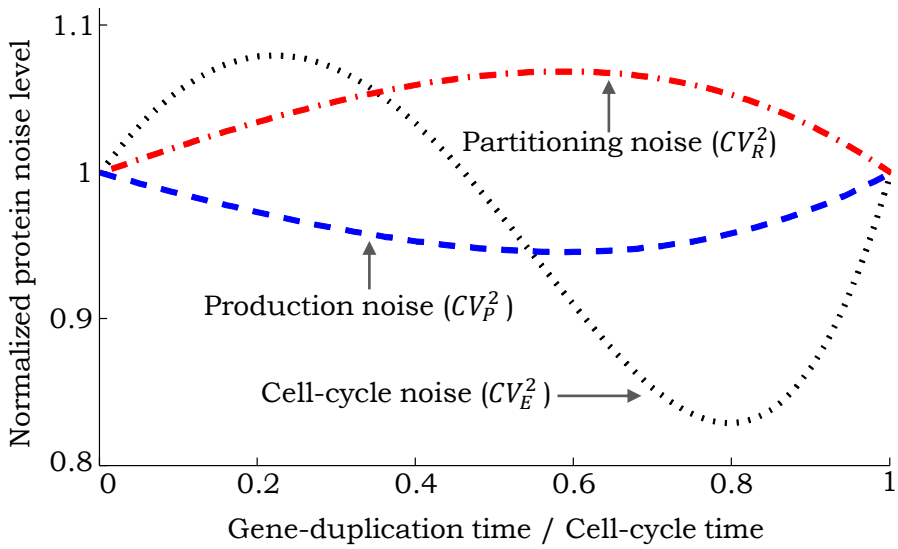

Figure 6. Contributions from different noise sources as a function of the timing of genome duplication for $CV_{T_1}^2 = CV_{T_2}^2 = 0$. Different noise components in (35) are plotted as a function of β, which represents the fraction of time within the cell-cycle at which gene-duplication occurs. The mean protein level is held constant by simultaneously changing the transcription rate k_x. Noise levels are normalized by their respective value at $\beta = 0$. The noise contribution from partitioning errors is maximized at $\beta \approx 0.6$. In contrast, the contribution from stochastic expression is minimum at $\beta \approx 0.6$. The extrinsic noise contribution from random gene-duplication and cell-division events is maximum at $\beta \approx 0.2$ and minimum at $\beta \approx 0.8$.

- For a non-bursty protein production process ($B = 1$) and binomial partitioning ($\alpha = 1$), the net noise from stochastic expression and partitioning is always $1/\overline{\langle x \rangle}$, the noise level predicted by a Poisson distribution.

We discuss our results on gene duplication in further detail and how noise formulas derived here can be used for estimating model parameters from single-cell expression data.

6.1 Affect of gene duplication on noise level

In this first-of-its-kind study, we have investigated how discrete two-fold changes in the transcription rate due do gene duplication affect the intercellular variability in protein levels. Not surprisingly, the timing of genome duplication has a strong effect on the mean protein level – $\overline{\langle x \rangle}$ changes by two-fold depending on whether the gene duplicates early ($\beta = 0$) or late ($\beta = 1$) in the cell-cycle. In contrast, the effect of β on noise is quite small. As β is varied keeping $\overline{\langle x \rangle}$ fixed, noise components deviate by $\approx 10\%$ from their values at $\beta = 0$ (Fig. 6). Recall that these results are for a stable protein, whose intracellular copy number accumulate in a bilinear fashion. A natural question to ask is how would these results change for an unstable protein?

Consider an unstable protein with half-life considerably shorter than the cell-cycle duration. This rapid turnover ensures that the protein level equilibrates instantaneously after cell-division and gene-duplication events. Let γ_x denote the protein decay rate. Then, the mean protein level before and after genome duplication is $\overline{\langle x \rangle} = k_x \langle B \rangle / \gamma_x$ and $\overline{\langle x \rangle} = 2k_x \langle B \rangle / \gamma_x$, respectively. Note that in the limit of large γ_x there is no noise contribution form partitioning errors since errors incurred at the time of cell division would be instantaneously corrected. The extrinsic noise, which can be interpreted as the protein noise

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

level for deterministic protein production and decay is obtained as (see Appendix F)

$$CV_E^2 = \frac{(1-\beta)\beta}{(2-\beta)^2}. \tag{37}$$

When $\beta = 0$ or 1, the transcription rate and the protein level are constant within the cell cycle and $CV_E^2 = 0$. Moreover, CV_E^2 is maximized at $\beta = 2/3$ with a value of 1/12. Thus, in contrast to a stable protein, extrinsic noise in an unstable protein is strongly dependent on the timing of gene duplication. Next, consider the intrinsic noise component. Analysis in Appendix F shows that the noise contribution from random protein production and decay is

$$CV_P^2 = \frac{1}{2}\left(\frac{\langle B^2\rangle}{\langle B\rangle} + 1\right)\frac{1}{\overline{\langle x\rangle}}, \quad \overline{\langle x\rangle} = \frac{k_x\langle B\rangle(2-\beta)}{\gamma_x}. \tag{38}$$

While the mean protein level is strongly dependent on β, the intrinsic noise Fano factor $= CV_P^2 \times \langle x\rangle$ is independent of it. Thus, similar to what was observed for a stable protein, the intrinsic noise in an unstable protein is invariant of β for a fixed $\overline{\langle x\rangle}$. Overall, these results suggest that studies quantifying intrinsic noise in gene expression models, or using intrinsic noise to estimate model parameters (see below) can ignore the effects of gene duplication. Finally, note that the mean and noise levels obtained for an unstable protein are independent of the cell-cycle time T.

6.2 Parameter inference from single-cell data

Simple models of bursty expression and decay predict the distribution of protein levels to be negative binomial (or gamma distributed in the continuous framework) [71,72]. These distributions are characterized by two parameter – the burst arrival rate k_x and the average burst size $\langle B\rangle$, which can be estimated from measured protein mean and noise levels. This method has been used for estimating k_x and $\langle B\rangle$ across different genes in *E. coli* [47,73]. Our detailed model that takes into account partitioning errors predicts (ignoring gene duplication effects)

$$\text{Intrinsic noise} = \frac{4\alpha}{3(3+CV_T^2)}\frac{1}{\overline{\langle x\rangle}} + \frac{3CV_T^2+5}{3(3+CV_T^2)}\frac{\langle B^2\rangle}{\langle B\rangle}\frac{1}{\overline{\langle x\rangle}}. \tag{39}$$

Using $CV_T^2 \ll 1$ and a geometrically distributed B [50,74–76], (39) reduces to

$$\text{Intrinsic noise} = \frac{4\alpha}{9}\frac{1}{\overline{\langle x\rangle}} + \frac{5}{9}\frac{1+2\langle B\rangle}{\overline{\langle x\rangle}}. \tag{40}$$

Given measurements of intrinsic noise and the mean protein level, $\langle B\rangle$ can be estimated from (40) assuming $\alpha = 1$ (i.e., binomial partitioning). Once $\langle B\rangle$ is known, k_x is obtained from the mean protein level given by (14). Since for many genes $\langle B\rangle \approx 0.5 - 5$ [47], the contribution of the first term in (40) is significant, and ignoring it could lead to overestimation of $\langle B\rangle$. Overestimation would be even more severe if α happen to be much higher than 1, as would be the case for proteins that form aggregates or multimers [33]. One approach to estimate both $\langle B\rangle$ and α is to measure intrinsic noise changes in response to perturbing $\langle B\rangle$ by, for example, changing the mRNA translation rate through mutations in the ribosomal-binding sites (RBS). Consider a hypothetical scenario where the Fano Factor (intrinsic noise times the mean level) is 6. Let mutations in the RBS reduces $\overline{\langle x\rangle}$ by 50%, implying a 50% reduction in $\langle B\rangle$. If the Fano factor changes from 6 to 4 due to this mutation, then $\langle B\rangle = 3.6$ and $\langle\alpha\rangle = 3.25$.

Our recent work has shown that higher-order statistics of protein levels (i.e., skewness and kurtosis) or transient changes in protein noise levels in response to blocking transcription provide additional information for discriminating between noise mechanisms [77,78]. Up till now these studies have ignored noise sources in the cell-cycle process. It remains to be seen if such methods can be used for separating the noise contributions of partitioning errors and stochastic expression to reliably estimate $\langle B\rangle$ and α.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6.3 Integrating cell size and promoter switching

An important limitation of our modeling approach is that it does not take into account the size of growing cells. Recent experimental studies have provided important insights into the regulatory mechanisms controlling cell size [79–81]. More specifically, studies in *E. coli* and yeast argue for an "adder" model, where cell-cycle timing is controlled so as to add a constant volume between cell birth and division [82–84]. Assuming exponential growth, this implies that the time taken to complete cell-cycle is negatively correlated with cell size at birth. In addition, cell size also affects gene expression – in mammalian cells transcription rates linearly increase with the cell size [85]. Thus, as cells become bigger they also produce more mRNAs to ensure gene product concentrations remains more or less constant. An important direction of future work would to explicitly include cell size with size-dependent expression and timing of cell division determined by the adder model. This formulation will for the first time, allow simultaneous investigation of stochasticity in cell size, protein molecular count and concentration.

Our study ignores genetic promoter switching between active and inactive states, which has been shown to be a major source of noise in the expression of genes across organisms [86–95]. Taking into account promote switching is particularly important for genome duplication studies, where doubling the number of gene copies could lead to more efficient averaging of promoter fluctuations. Another direction of future work will be to incorporate this addition noise source into the modeling framework and investigate its contribution as a function of gene-duplication timing.

Appendix

A Mean of protein in the presence of cell-cycle variations

Based on standard stochastic formulation of chemical kinetics [96, 97], the model introduced in Figure 2A coupled with phase-type distribution introduced in Figure 3 contains the following stochastic events

Event	Reset	Propensity
Protein production	$x(t) \mapsto x(t)+u$	$k_x p''_u$
Phase-type Evolution	$g_{ij}(t) \mapsto g_{ij}(t)-1,$ $g_{i(j+1)}(t) \mapsto g_{i(j+1)}(t)+1$	$kg_{ij},$ $i \in \{2,\ldots,n\},$ $j \in \{1,\ldots,i-1\}$
Cell-division	$x(t_s) \mapsto x_+(t_s),$ $g_{jj}(t_s) \mapsto 0,$ $g_{i1}(t_s) \mapsto g_{i1}(t_s)+1$	$kp_i \sum_{j=1}^{n} g_{jj},$ $i \in \{1,\ldots,n\}$

Note that $x_+(t_s)$ is protein level after division, characteristics of $x_+(t_s)$ is related to protein level before division as shown in equation (5) of the main text. Whenever an event occurs, protein level and states of phase-type distribution change based on the stoichiometries shown in the second column of the table. The third column of table shows event propensity function $f(x, g_{ij})$, which determines how often reactions occur, i.e., the probability that an event occurs in the next infinitesimal time interval $(t, t+dt]$ is $f(x, g_{ij})dt$. Protein production is a stochastic event which happens in bursts, each burst generates B molecules where

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B is a general random variable with distribution

$$\text{Probability}\{B = u\} = p''_u, \quad u \in \{0, 1, \ldots, \infty\}. \tag{41}$$

The probability of having a burst in the time interval $(t, t+dt]$ is $k_x p''_u dt$. Events related to time evolution of phase-type distribution happen with a constant rate k. Cell-division changes both the level of protein and states of phase-type. This event contains start of new cell-cycle, hence whenever this event occurs, the last state of phase-type distribution resets to zero, and a new cell-cycle which is sum of i exponentials starts with probability p_i; protein count level also resets to $x_+(t_s)$. The probability of cell-division and starting a new cell-cycle from state g_{i1} in the time interval $(t, t+dt]$ is $kp_i \sum_{j=1}^{n} dt$.

Theorem 1 of [55] gives the time derivative of the expected value of any function $\varphi(x, g_{ij})$ as

$$\frac{d\langle\varphi(x, g_{ij})\rangle}{dt} = \left\langle \sum_{Events} \Delta\varphi(x, g_{ij}) \times f(x, g_{ij}) \right\rangle, \tag{42}$$

where $\Delta\varphi(x, g_{ij})$ is a change in φ when an event occurs. Based on this setup, mean dynamics of protein can be written by choosing φ to be x

$$\begin{aligned} \frac{d\langle x\rangle}{dt} &= k_x\langle B\rangle + k\left\langle \sum_{j=1}^{n} (\frac{x}{2} - x) g_{jj} \right\rangle \Rightarrow \\ \frac{d\langle x\rangle}{dt} &= k_x\langle B\rangle - \frac{k}{2}\left\langle \sum_{j=1}^{n} x g_{jj} \right\rangle, \end{aligned} \tag{43}$$

where we replaced conditional expected value of x_+ by $x/2$ based on relation between statistical properties of x_+ and x shown in equation (5).

Dynamics of $\langle x\rangle$ is not closed and depends to moments $\langle x g_{jj}\rangle$, hence in order to have a closed set of equations we add new moments dynamics by selecting φ to be $x g_{ij}$. We do it in two steps: first we write the moment dynamics of $\langle x g_{11}\rangle$

$$\frac{d\langle x g_{11}\rangle}{dt} = k_x\langle B\rangle\langle g_{11}\rangle + \frac{k}{2} p_1 \left\langle x g_{11}^2 \right\rangle - k p_1 \left\langle x g_{11}^2 \right\rangle - k \sum_{i=2}^{n} p_i \left\langle x g_{11} \right\rangle. \tag{44}$$

In the equation (9) of the main text it has been shown that

$$\langle g_{ij}^n x^m\rangle = \langle g_{ij} x^m\rangle, \quad n \in \{1, 2, \ldots\}, \tag{45}$$

thus the term $\left\langle x g_{11}^2 \right\rangle$ will simplify as

$$\left\langle x g_{11}^2 \right\rangle = \left\langle x g_{11} \right\rangle, \tag{46}$$

and the dynamics of $\langle x g_{11}\rangle$ can be written as

$$\frac{d\langle x g_{11}\rangle}{dt} = k_x\langle B\rangle\langle g_{11}\rangle + \frac{k}{2} p_1 \left\langle x g_{11} \right\rangle - k \left\langle x g_{11} \right\rangle. \tag{47}$$

In the second step we write dynamics of the moments of the form $\langle x g_{ij}\rangle$ other than $\langle x g_{11}\rangle$

$$\frac{d\langle x g_{i1}\rangle}{dt} = k_x\langle B\rangle\langle g_{i1}\rangle + k p_i \left\langle \sum_{j=1}^{n} (\frac{x}{2} + \frac{x}{2} g_{i1} - x g_{i1}) g_{jj} \right\rangle - k\langle x g_{i1}\rangle, \tag{48a}$$

$$\frac{d\langle x g_{ij}\rangle}{dt} = k_x\langle B\rangle\langle g_{ij}\rangle - k\langle x g_{ij}\rangle + k\langle x g_{i(j-1)}\rangle, \quad j \in \{2, \ldots, i\}, \tag{48b}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where dynamics of $\langle x g_{i1} \rangle$ can be written as

$$\frac{d\langle x g_{i1}\rangle}{dt} = \frac{k_x \langle B \rangle p_i}{i} + k p_i \left\langle \sum_{j=1}^{n} \frac{x}{2} g_{jj} \right\rangle + k p_i \left\langle \sum_{j=1}^{n} -\frac{x}{2} g_{i1} g_{jj} \right\rangle - k \langle x g_{i1} \rangle. \tag{49}$$

The equation (10) in the main text shows that

$$\langle g_{ij} g_{rq} x^m \rangle = 0, \text{ if } i \neq r \text{ or } j \neq q, \tag{50}$$

hence $\left\langle \sum_{j=1}^{n} \frac{x}{2} g_{i1} g_{jj} \right\rangle = 0$, and equation (49) simplifies to

$$\frac{d\langle x g_{i1}\rangle}{dt} = k_x \langle B \rangle \langle g_{i1} \rangle + \frac{k}{2} p_i \left\langle \sum_{j=1}^{n} x g_{jj} \right\rangle - k \langle x g_{i1} \rangle. \tag{51}$$

Further based on Figure 3 in the main text the probability of selecting a branch of i exponentials is p_i, and because all the transitions happen with a constant rate k, hence mean of each of these i states is

$$\langle g_{ij} \rangle = \frac{p_i}{i}. \tag{52}$$

Thus equations (47), (48b), and (51) can be compactly written as shown in equation (11).

B Moment dynamics of hybrid model introduced in Figure 2B

Stochastic hybrid system introduced in Figure 2B coupled with phase-type distribution contains the following stochastic events

Event	Reset	Propensity
Phase-type Evolution	$g_{ij}(t) \mapsto g_{ij}(t) - 1$, $g_{i(j+1)}(t) \mapsto g_{i(j+1)}(t) + 1$	$k g_{ij}$, $i \in \{2, \ldots, n\}$, $j \in \{1, \ldots, i-1\}$
Cell-division	$x(t_s) \mapsto x(t_s)/2$, $g_{jj}(t_s) \mapsto 0$, $g_{i1}(t_s) \mapsto g_{i1}(t_s) + 1$	$k p_i \sum_{j=1}^{n} g_{jj}$, $i \in \{1, \ldots, n\}$

and deterministic protein production dynamics

$$\dot{x} = k_x \langle B \rangle. \tag{53}$$

Time derivative of the expected value of any function $\varphi(x, g_{ij})$ for this hybrid system can be written as [55]

$$\frac{d\langle \varphi(x, g_{ij}) \rangle}{dt} = \left\langle \sum_{Events} \Delta\varphi(x, g_{ij}) \times f(x, g_{ij}) \right\rangle + \left\langle \frac{\partial \varphi(x, g_{ij})}{\partial x} k_x \langle B \rangle \right\rangle, \tag{54}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where the first term in the right-hand side is contributed from stochastic events and the second term is contributed from deterministic protein production dynamics. Based on this equation, the mean dynamics of the protein is calculated by choosing φ to be x

$$\frac{d\langle x\rangle}{dt} = k_x\langle B\rangle - \frac{k}{2}\left\langle \sum_{j=1}^{n} x g_{jj} \right\rangle, \tag{55}$$

which is the same as equation (43). In addition to mean, dynamics of $\langle x g_{ij}\rangle$ are also equal to their equation in the previous section.

The second order moment dynamics of protein can be expressed by choosing φ to be x^2

$$\frac{d\langle x^2\rangle}{dt} = 2k_x\langle B\rangle\langle x\rangle + k\left\langle \sum_{j=1}^{n}\left(\left(\frac{x}{2}\right)^2 - x^2\right) g_{jj} \right\rangle, \tag{56}$$

which can be simplified as

$$\frac{d\langle x^2\rangle}{dt} = 2k_x\langle B\rangle\langle x\rangle - \frac{3k}{4}\left\langle \sum_{j=1}^{n} x^2 g_{jj} \right\rangle. \tag{57}$$

In order to have a closed set of equations we select φ to be of the form $x^2 g_{ij}$. At the first step we write moment dynamics of $\langle x^2 g_{11}\rangle$

$$\frac{d\langle x^2 g_{11}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{11}\rangle + \frac{k}{4}p_1\left\langle x^2 g_{11}^2\right\rangle - kp_1\left\langle x^2 g_{11}^2\right\rangle - k\sum_{i=2}^{n} p_i\langle x^2 g_{11}\rangle. \tag{58}$$

Based on equation (9) of the main text, the term $\left\langle x^2 g_{11}^2\right\rangle$ simplifies as

$$\left\langle x^2 g_{11}^2\right\rangle = \left\langle x^2 g_{11}\right\rangle, \tag{59}$$

hence dynamics of $\langle x^2 g_{11}\rangle$ will be

$$\frac{d\langle x^2 g_{11}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{11}\rangle + \frac{k}{4}p_1\left\langle x^2 g_{11}\right\rangle - k\langle x^2 g_{11}\rangle. \tag{60}$$

In the second step, we write dynamics of moments $\langle x^2 g_{ij}\rangle$ when $g_{ij} \neq g_{11}$

$$\frac{d\langle x^2 g_{i1}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{i1}\rangle + kp_i\left\langle \sum_{j=1}^{n}\left(\frac{x^2}{4} + \frac{x^2}{4}g_{i1} - x^2 g_{i1}\right) g_{jj} \right\rangle - k\langle x^2 g_{i1}\rangle, \tag{61a}$$

$$\frac{d\langle x^2 g_{ij}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{ij}\rangle - k\langle x^2 g_{ij}\rangle + k\langle x^2 g_{(i-1)j}\rangle,\ j = \{2,\ldots,i\}, \tag{61b}$$

where dynamics of $\langle x^2 g_{i1}\rangle$ can be shown to follow

$$\frac{d\langle x^2 g_{i1}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{i1}\rangle + \frac{k}{4}p_i\left\langle \sum_{j=1}^{n} x^2 g_{jj} \right\rangle - \frac{3k}{4}p_i\left\langle \sum_{j=1}^{n} x^2 g_{i1} g_{jj} \right\rangle - k\langle x^2 g_{i1}\rangle. \tag{62}$$

Based on equation (10) in the main text $\left\langle \sum_{j=1}^{n} x^2 g_{i1} g_{jj} \right\rangle = 0$, thus equation (62) simplifies to

$$\frac{d\langle x^2 g_{i1}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{i1}\rangle + \frac{k}{4}p_i\left\langle \sum_{j=1}^{n} x^2 g_{jj} \right\rangle - k\langle x^2 g_{i1}\rangle. \tag{63}$$

Equations (60), (61b), and (63) can be compactly written as equation (16) in the main text.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

C Moment dynamics of hybrid model introduced in Figure 2C

Stochastic hybrid system introduced in Figure 2C coupled with phase-type distribution contains the following stochastic events

Event	Reset	Propensity
Phase-type Evolution	$g_{ij}(t) \mapsto g_{ij}(t)-1,$ $g_{i(j+1)}(t) \mapsto g_{i(j+1)}(t)+1$	$kg_{ij},$ $i \in \{2,\ldots,n\},$ $j \in \{1,\ldots,i-1\}$
Cell-division	$x(t_s) \mapsto x_+(t_s),$ $g_{jj}(t_s) \mapsto 0,$ $g_{i1}(t_s) \mapsto g_{i1}(t_s)+1$	$kp_i \sum_{j=1}^{n} g_{jj},$ $i \in \{1,\ldots,n\}$

and deterministic protein production dynamics

$$\dot{x} = k_x \langle B \rangle. \tag{64}$$

Note that in this model $x(t)$ is a continuous random variable, thus we also use a continuous distribution to describe $x_+(t_s)$, however statistical properties of $x_+(t_s)$ is still given by (5). For this model we still can use equation (54) to derive moment dynamics; equations describing time evolution of mean and $\langle x g_{ij} \rangle$ are the same as previous models, thus mean of protein for this model is equal to its value in Appendix A. The second order moment dynamics of protein can be written by choosing φ to be x^2 in equation (54)

$$\frac{d\langle x^2 \rangle}{dt} = 2k_x \langle B \rangle \langle x \rangle + k \left\langle \sum_{j=1}^{n} \left(\frac{x^2}{4} + \frac{\alpha x}{4} - x^2 \right) g_{jj} \right\rangle, \tag{65}$$

where conditional expected value of x_+^2 is substituted based on equation (5). Dynamics of $\langle x^2 \rangle$ can be simplified as

$$\frac{d\langle x^2 \rangle}{dt} = 2k_x \langle B \rangle \langle x \rangle + \frac{\alpha k}{4} \left\langle \sum_{j=1}^{n} x g_{jj} \right\rangle - \frac{3k}{4} \left\langle \sum_{j=1}^{n} x^2 g_{jj} \right\rangle. \tag{66}$$

The same as before we add dynamics of the form $\langle x^2 g_{ij} \rangle$ to have a closed set of dynamics. First we add dynamics of $\langle x^2 g_{11} \rangle$

$$\frac{d\langle x^2 g_{11} \rangle}{dt} = 2k_x \langle B \rangle \langle x g_{11} \rangle + \frac{\alpha k}{4} p_1 \langle x g_{11}^2 \rangle + \frac{k}{4} p_1 \langle x^2 g_{11}^2 \rangle - k p_1 \langle x^2 g_{11}^2 \rangle - k \sum_{i=2}^{n} p_i \langle x^2 g_{11} \rangle, \tag{67}$$

Based on equation (9) of the main text dynamics of $\langle x^2 g_{11} \rangle$ simplifies to

$$\frac{d\langle x^2 g_{11} \rangle}{dt} = 2k_x \langle B \rangle \langle x g_{11} \rangle + \frac{\alpha k}{4} p_1 \langle x g_{11} \rangle + \frac{k}{4} p_1 \langle x^2 g_{11} \rangle - k \langle x^2 g_{11} \rangle. \tag{68}$$

Now we express dynamics of moments $\langle x^2 g_{ij} \rangle$ for $g_{ij} \neq g_{11}$

$$\frac{d\langle x^2 g_{i1} \rangle}{dt} = 2k_x \langle B \rangle \langle x g_{i1} \rangle + k p_i \left\langle \sum_{j=1}^{n} \left(\frac{x^2}{4} + \frac{x^2}{4} g_{i1} + \frac{\alpha x}{4} + \frac{\alpha x}{4} g_{i1} - x^2 g_{i1} \right) g_{jj} \right\rangle - k \langle x^2 g_{i1} \rangle, \tag{69a}$$

$$\frac{d\langle x^2 g_{ij} \rangle}{dt} = 2k_x \langle B \rangle \langle x g_{ij} \rangle - k \langle x^2 g_{ij} \rangle + k \langle x^2 g_{(i-1)j} \rangle, \; j = \{2, \ldots, i\}, \tag{69b}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where dynamics of $\langle x^2 g_{i1}\rangle$ can be shown as

$$\begin{aligned}\frac{d\langle x^2 g_{i1}\rangle}{dt} =& 2k_x\langle B\rangle\langle x g_{i1}\rangle + \frac{\alpha k}{4}p_i\left\langle\sum_{j=1}^{n} x g_{jj}\right\rangle + \frac{k}{4}p_i\left\langle\sum_{j=1}^{n} x^2 g_{jj}\right\rangle \\ &+ \frac{\alpha k}{4}p_i\left\langle\sum_{j=1}^{n} x g_{i1}g_{jj}\right\rangle - \frac{3k}{4}p_i\left\langle\sum_{j=1}^{n} x^2 g_{i1}g_{jj}\right\rangle - k\langle x^2 g_{i1}\rangle.\end{aligned} \tag{70}$$

Based on equation (10) in the main text $\left\langle\sum_{j=1}^{n} x^2 g_{i1}g_{jj}\right\rangle = 0$, and $\left\langle\sum_{j=1}^{n} x g_{i1}g_{jj}\right\rangle = 0$, hence equation (70) simplifies to

$$\frac{d\langle x^2 g_{i1}\rangle}{dt} = 2k_x\langle B\rangle\langle x g_{i1}\rangle + \frac{\alpha k}{4}p_i\left\langle\sum_{j=1}^{n} x g_{jj}\right\rangle + \frac{k}{4}p_i\left\langle\sum_{j=1}^{n} x^2 g_{jj}\right\rangle - k\langle x^2 g_{i1}\rangle. \tag{71}$$

Equations (66), (68), (69b), and (71) can be compactly written as equation (23) in the main text.

D Second and third-order moment dynamics of the full model

Based on model introduced in Appendix A, second order moment dynamics of protein is expressed by choosing φ to be x^2 in equation (42),

$$\frac{d\langle x^2\rangle}{dt} = k_x\langle B^2\rangle + 2k_x\langle B\rangle\langle x\rangle + k\left\langle\sum_{j=1}^{n}\left(\frac{x^2}{4} + \frac{\alpha x}{4} - x^2\right)g_{jj}\right\rangle, \tag{72}$$

where conditional expected value of x_+^2 is substituted based on equation (5). Dynamics of $\langle x^2\rangle$ can be simplified as

$$\frac{d\langle x^2\rangle}{dt} = k_x\langle B^2\rangle + 2k_x\langle B\rangle\langle x\rangle + \frac{\alpha k}{4}\left\langle\sum_{j=1}^{n} x g_{jj}\right\rangle - \frac{3k}{4}\left\langle\sum_{j=1}^{n} x^2 g_{jj}\right\rangle. \tag{73}$$

The same as before we add dynamics of the form $\langle x^2 g_{ij}\rangle$ to have a closed set of moments. First we write dynamics of $\langle x^2 g_{11}\rangle$

$$\frac{d\langle x^2 g_{11}\rangle}{dt} = k_x\langle B^2\rangle p_1 + 2k_x\langle B\rangle\langle x g_{11}\rangle + \frac{\alpha k}{4}p_1\left\langle x g_{11}^2\right\rangle + \frac{k}{4}p_1\left\langle x^2 g_{11}^2\right\rangle - kp_1\left\langle x^2 g_{11}^2\right\rangle - k\sum_{i=2}^{n} p_i\langle x^2 g_{11}\rangle, \tag{74}$$

Based on equation (9) of the main text dynamics of $\langle x^2 g_{11}\rangle$ simplifies to

$$\frac{d\langle x^2 g_{11}\rangle}{dt} = k_x\langle B^2\rangle p_1 + 2k_x\langle B\rangle\langle x g_{11}\rangle + \frac{\alpha k}{4}p_1\left\langle x g_{11}\right\rangle + \frac{k}{4}p_1\left\langle x^2 g_{11}\right\rangle - k\langle x^2 g_{11}\rangle. \tag{75}$$

Next, dynamics of moments $\langle x^2 g_{ij}\rangle$ when $g_{ij} \neq g_{11}$ can be written as

$$\begin{aligned}\frac{d\langle x^2 g_{i1}\rangle}{dt} =& \frac{k_x\langle B^2\rangle p_i}{i} + 2k_x\langle B\rangle\langle x g_{i1}\rangle \\ &+ kp_i\left\langle\sum_{j=1}^{n}\left(\frac{x^2}{4} + \frac{x^2}{4}g_{i1} + \frac{\alpha x}{4} + \frac{\alpha x}{4}g_{i1} - x^2 g_{i1}\right)g_{jj}\right\rangle - k\langle x^2 g_{i1}\rangle,\end{aligned} \tag{76a}$$

$$\frac{d\langle x^2 g_{ij}\rangle}{dt} = \frac{k_x\langle B^2\rangle p_i}{i} + 2k_x\langle B\rangle\langle x g_{ij}\rangle - k\langle x^2 g_{ij}\rangle + k\langle x^2 g_{(i-1)j}\rangle,\ j = \{2,\ldots,i\}, \tag{76b}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where dynamics of $\langle x^2 g_{i1}\rangle$ can be shown as

$$\begin{aligned}\frac{d\langle x^2 g_{i1}\rangle}{dt} =& \frac{k_x\langle B^2\rangle p_i}{i} + 2k_x\langle B\rangle\langle x g_{i1}\rangle + \frac{\alpha k}{4}p_i\left\langle\sum_{j=1}^{n} x g_{jj}\right\rangle + \frac{k}{4}p_i\left\langle\sum_{j=1}^{n} x^2 g_{jj}\right\rangle \\ &+ \frac{\alpha k}{4}p_i\left\langle\sum_{j=1}^{n} x g_{i1} g_{jj}\right\rangle - \frac{3k}{4}p_i\left\langle\sum_{j=1}^{n} x^2 g_{i1} g_{jj}\right\rangle - k\langle x^2 g_{i1}\rangle.\end{aligned} \tag{77}$$

Based on equation (10) in the main text $\left\langle\sum_{j=1}^{n} x^2 g_{i1} g_{jj}\right\rangle = 0$ and $\left\langle\sum_{j=1}^{n} x g_{i1} g_{jj}\right\rangle = 0$, hence equation (77) simplifies to

$$\frac{d\langle x^2 g_{i1}\rangle}{dt} = \frac{k_x\langle B^2\rangle p_i}{i} + 2k_x\langle B\rangle\langle x g_{i1}\rangle + \frac{\alpha k}{4}p_i\left\langle\sum_{j=1}^{n} x g_{jj}\right\rangle + \frac{k}{4}p_i\left\langle\sum_{j=1}^{n} x^2 g_{jj}\right\rangle - k\langle x^2 g_{i1}\rangle. \tag{78}$$

Equations (73), (75), (76b), and (78) can be compactly written as equation (26) in the main text.

E Contribution of different sources of stochasticity in protein by taking into account gene-duplication

We study the contribution of different sources of stochasticity by using models introduced in Figure S1. The cell-cycle time consists of two time intervals: the time interval before gene-duplication and the time after gene-duplication. These time intervals are modeled by using two independent phase-type distributions as shown in Figure S2. Based on phase-type characteristics mean of the states of the first phase-type $\langle s_{ij}\rangle$ and the second phase-type $\langle g_{ij}\rangle$ are

$$\begin{aligned}\langle s_{ij}\rangle &= \frac{p_i}{i}\beta, \quad i\in\{1,\dots,n_1\}, \quad j\in\{1,\dots,i\}, \\ \langle g_{ij}\rangle &= \frac{p'_i}{i}(1-\beta), \quad i\in\{1,\dots,n_2\}, \quad j\in\{1,\dots,i\},\end{aligned} \tag{79}$$

where β is defined as

$$\beta := \frac{\text{Mean time interval before gene-duplication}}{\text{Mean cell-cycle time}} = \frac{\langle T_1\rangle}{\langle T\rangle}. \tag{80}$$

We start our analysis by deriving mean level of protein in the next section.

E.1 Mean of protein count level in the presence of gene-duplication

After gene-duplication the amount of genes expressing a specific protein doubles. Thus the rate of protein production increases by a factor of two as shown in Figure S1A. This model coupled with phase-type distributions contains the following stochastic events

Note that in the protein production event, before gene-duplication all the states g_{ij} are zero thus propensity function will be $k_x p''_u$. After gene-duplication and before division, one of the states g_{ij} is one hence propensity function will be $2k_x p''_u$. In time of gene-duplication, states of the first phase-type will reset to zero and state g_{i1} of the second distribution will be selected with probability p'_i; hence propensity function of gene-duplication event is $k_1 p'_i \sum_{j=1}^{n_1} s_{jj}$. At the end of cell-cycle, states of the second phase-type will reset to zero and a new cell-cycle which is sum of i exponentials will be selected with probability p_i; thus propensity function of cell-division event is $k_2 p_i \sum_{j=1}^{n_1} g_{jj}$.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

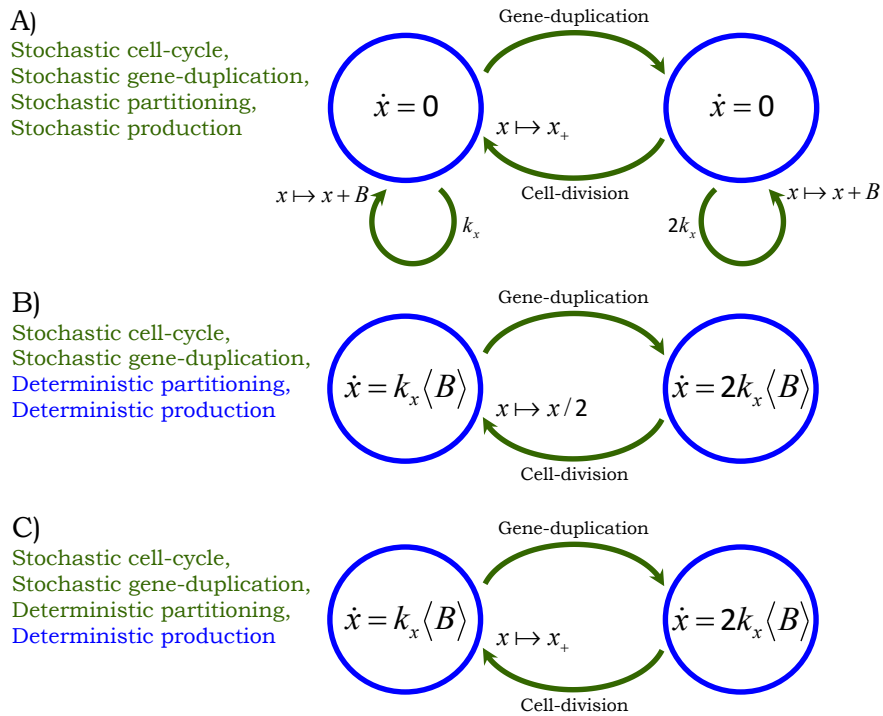

Figure S1: **Stochastic hybrid models for quantifying different sources of noise.** Gene-duplication and cell-division times are random events. **A)** Protein production happens in random bursts with burst frequency k_x. After gene-duplication event burst frequency doubles ($2k_x$). In the time of division proteins will be distributed between mother and daughter cells randomly, and the protein burst frequency will be k_x again. **B)** Protein production is considered in a deterministic fashion, and after gene-duplication dynamics of protein production is multiplied by a factor of two, i.e., $\dot{x} = 2k_x\langle B\rangle$. In the division event proteins are distributed between mother and daughter cells equally. Thus the only stochastic events are duplication and division events. **C)** Protein is produced in a deterministic fashion, but in time of division protein levels in daughter and mother cells are random. Thus duplication, division, and partitioning are random events.

Theorem 1 of [55] gives the time derivative of the expected value of any function $\varphi(x, s_{ij}, g_{ij})$ as

$$\frac{d\langle\varphi(x, s_{ij}, g_{ij})\rangle}{dt} = \left\langle \sum_{Events} \Delta\varphi(x, s_{ij}, g_{ij}) \times f(x, s_{ij}, g_{ij}) \right\rangle, \tag{81}$$

where $\Delta\varphi(x, s_{ij}, g_{ij})$ is a change in φ when an event occurs. The first-order moment dynamic of this model can be expressed by selecting φ to be x in equation (81)

$$\frac{d\langle x\rangle}{dt} = k_x\langle B\rangle \left(1 + \left\langle \sum_{i=1}^{n_2}\sum_{j=1}^{i} g_{ij} \right\rangle\right) - k_2 \left\langle \sum_{j=1}^{n_2} (\frac{x}{2} - x) g_{jj} \right\rangle, \tag{82}$$

where conditional expected value of x_+ is replaced from equation (5); by using equation (79) mean

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

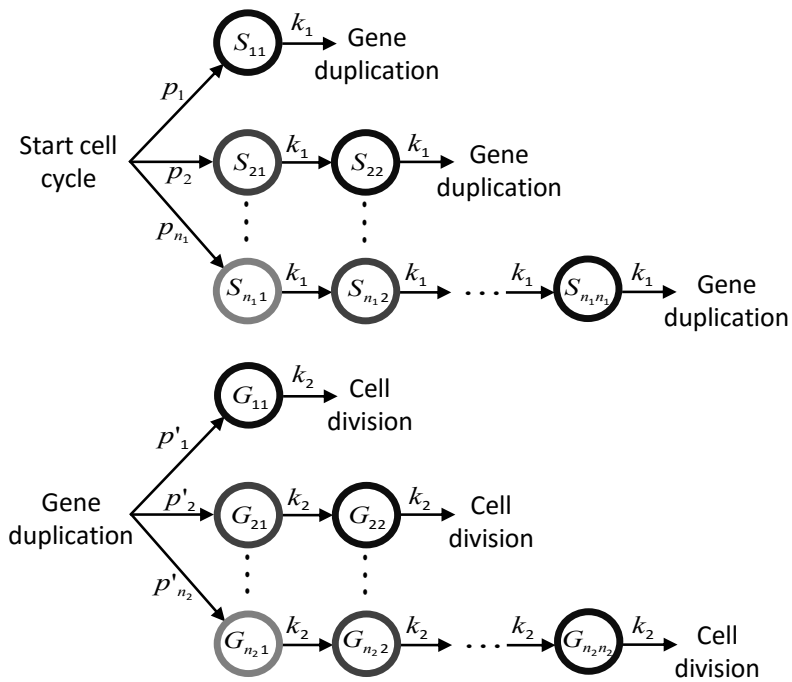

Figure S2: **Cell-cycle time consists of two time intervals: at the end of the first interval gene duplicates, and at the end of the second one cell divides.** Two independent phase-type distributions are used to model cell-cycle time in the presence of genome duplication. The states of the first distribution are denoted by $S_{ij},\ i=\{1,\dots,n_1\},\ j=\{1,\dots,i\}$; transition between these states happens at a constant rate k_1. The states of the second distribution are shown by $G_{ij},\ i=\{1,\dots,n_2\},\ j=\{1,\dots,i\}$, and transition between these states occurs at a rate k_2.

dynamics can be simplified as

$$\frac{d\langle x\rangle}{dt} = k_x\langle B\rangle\,(2-\beta) - \frac{k_2}{2}\left\langle \sum_{j=1}^{n_2} x g_{jj}\right\rangle, \tag{83}$$

Mean dynamics is not closed thus we add dynamics of $\langle x s_{ij}\rangle,\ i=\{1,\dots,n_1\},\ j=\{1,\dots,i\}$ and $\langle x g_{ij}\rangle,\ i=\{1,\dots,n_1\},\ j=\{1,\dots,i\}$ to have a closed set of moment equations. These moment dynamics are simplified by using equations (5), (9), (10) and (79) as

$$\frac{d\langle x s_{i1}\rangle}{dt} = \frac{k_x\langle B\rangle p_i'\beta}{i} + \frac{k_2}{2}p_i'\left\langle \sum_{j=1}^{n_2} x g_{jj}\right\rangle - k_1\langle x s_{i1}\rangle, \tag{84a}$$

$$\frac{d\langle x s_{ij}\rangle}{dt} = \frac{k_x\langle B\rangle p_i'\beta}{i} - k_1\langle x s_{ij}\rangle + k_1\langle x s_{i(j-1)}\rangle,\quad j=\{2,\dots,i\}, \tag{84b}$$

$$\frac{d\langle x g_{i1}\rangle}{dt} = \frac{2k_x\langle B\rangle p_i(1-\beta)}{i} + k_1 p_i\left\langle \sum_{j=1}^{n_1} x s_{jj}\right\rangle - k_2\langle x g_{i1}\rangle, \tag{84c}$$

$$\frac{d\langle x g_{ij}\rangle}{dt} = \frac{2k_x\langle B\rangle p_i(1-\beta)}{i} - k_2\langle x g_{ij}\rangle + k_2\langle x g_{i(j-1)}\rangle,\ j=\{2,\dots,i\}. \tag{84d}$$

In order to find the mean of protein, first we need to find the moments $\overline{\langle x s_{ij}\rangle},\ i=\{1,\dots,n_1\},\ j=$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Event	Reset	Propensity
Protein production	$x(t) \mapsto x(t)+u$	$k_x p''_u \left(1+\sum_{i=1}^{n_2}\sum_{j=1}^{i} g_{ij}\right)$
First phase-type evolution	$s_{ij}(t) \mapsto s_{ij}(t)-1,$ $s_{i(j+1)}(t) \mapsto s_{i(j+1)}(t)+1$	$k_1 s_{ij},$ $i \in \{2,\ldots,n_1\},$ $j \in \{1,\ldots,i-1\}$
Gene-duplication	$s_{jj}(t) \mapsto 0,$ $g_{i1}(t) \mapsto g_{i1}(t)+1$	$k_1 p'_i \sum_{j=1}^{n_1} s_{jj},$ $i \in \{1,\ldots,n_2\}$
Second phase-type evolution	$g_{ij}(t) \mapsto g_{ij}(t)-1,$ $g_{i(j+1)}(t) \mapsto g_{i(j+1)}(t)+1$	$k_2 g_{ij},$ $i \in \{2,\ldots,n_2\},$ $j \in \{1,\ldots,i-1\}$
Cell-division	$x(t_s) \mapsto x_+(t_s),$ $g_{jj}(t_s) \mapsto 0,$ $s_{i1}(t_s) \mapsto s_{i1}(t_s)+1$	$k_2 p_i \sum_{j=1}^{n_2} g_{jj},$ $i \in \{1,\ldots,n_1\}$

$\{1,\ldots,i\}$ and $\overline{\langle xg_{ij}\rangle}$, $i=\{1,\ldots,n_2\}$, $j=\{1,\ldots,i\}$. For calculating these moments we should calculate the term $\overline{\left\langle \sum_{j=1}^{n_2} xg_{jj}\right\rangle}$; this term can be obtained by analyzing equation (83) in steady-state

$$k_x\langle B\rangle(2-\beta) = \frac{k_2}{2}\overline{\left\langle \sum_{j=1}^{n_2} xg_{jj}\right\rangle} \Rightarrow \overline{\left\langle \sum_{j=1}^{n_2} xg_{jj}\right\rangle} = \frac{2k_x\langle B\rangle(2-\beta)}{k_2}. \tag{85}$$

By having this term, we calculate $\overline{\langle xs_{ij}\rangle}$ by recursion process: we start by calculating $\overline{\langle xs_{i1}\rangle}$ by substituting equation (85) in equation (84a). In the next step we use the definition we derived for $\overline{\langle xs_{i1}\rangle}$ to calculate $\overline{\langle xs_{i2}\rangle}$ from equation (84b). We continue this process until we derive all the moments

$$\overline{\langle xs_{ij}\rangle} = \frac{k_x\langle B\rangle}{k_1} p'_i \left(\beta\frac{j}{i} + (2-\beta)\right),\ i=\{1,\ldots,n_1\},\ j=\{1,\ldots,i\}. \tag{86}$$

Now we need to calculate the moments $\overline{\langle xg_{ij}\rangle}$, $i=\{1,\ldots,n_2\}$, $j=\{1,\ldots,i\}$, thus we need the expression of the term $\overline{\left\langle \sum_{j=1}^{n_1} xs_{jj}\right\rangle}$; from equation (86) we have the following

$$\overline{\left\langle \sum_{j=1}^{n_1} xs_{jj}\right\rangle} = \frac{2k_x\langle B\rangle}{k_1}. \tag{87}$$

Substituting this term in equations (84c) and (84d) result in

$$\overline{\langle xg_{ij}\rangle} = \frac{2k_x\langle B\rangle}{k_2} p_i \left((1-\beta)\frac{j}{i} + 1\right),\ i=\{1,\ldots,n_2\},\ j=\{1,\ldots,i\}. \tag{88}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Note that

$$\begin{aligned}\sum_{i=1}^{n_1}\sum_{j=1}^{i} s_{ij} + \sum_{i=1}^{n_2}\sum_{j=1}^{i} g_{ij} = 1 \Rightarrow \langle x\rangle &= \left\langle x\left(\sum_{i=1}^{n_1}\sum_{j=1}^{i} s_{ij} + \sum_{i=1}^{n_2}\sum_{j=1}^{i} g_{ij}\right)\right\rangle \\ \Rightarrow \overline{\langle x\rangle} &= \sum_{i=1}^{n_1}\sum_{j=1}^{i}\overline{\langle x s_{ij}\rangle} + \sum_{i=1}^{n_2}\sum_{j=1}^{i}\overline{\langle x g_{ij}\rangle}.\end{aligned} \tag{89}$$

Thus by adding all the term calculated here and using equation (7) mean of protein can be calculated as

$$\overline{\langle x\rangle} = \frac{k_x\langle B\rangle\langle T_1\rangle\left(4-\beta+\beta CV_{T_1}^2\right)}{2} + k_x\langle B\rangle\langle T_2\rangle\left(3-\beta+(1-\beta)CV_{T_2}^2\right). \tag{90}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

E.2 Noise in protein count level contributed from cell-cycle time

In order to calculate the noise contributed from cell-cycle time variation, the model introduced in Figure S1B coupled with phase-type distributions is used. This model contains following stochastic events

Event	Reset	Propensity
First phase-type evolution	$s_{ij}(t) \mapsto s_{ij}(t)-1,$ $s_{i(j+1)}(t) \mapsto s_{i(j+1)}(t)+1$	$k_1 s_{ij},$ $i \in \{2,\ldots,n_1\},$ $j \in \{1,\ldots,i-1\}$
Gene-duplication	$s_{jj}(t) \mapsto 0,$ $g_{i1}(t) \mapsto g_{i1}(t)+1$	$k_1 p'_i \sum_{j=1}^{n_1} s_{jj},$ $i \in \{1,\ldots,n_2\}$
Second phase-type evolution	$g_{ij}(t) \mapsto g_{ij}(t)-1,$ $g_{i(j+1)}(t) \mapsto g_{i(j+1)}(t)+1$	$k_2 g_{ij},$ $i \in \{2,\ldots,n_2\},$ $j \in \{1,\ldots,i-1\}$
Cell-division	$x(t_s) \mapsto x(t_s)/2,$ $g_{jj}(t_s) \mapsto 0,$ $s_{i1}(t_s) \mapsto s_{i1}(t_s)+1$	$k_2 p_i \sum_{j=1}^{n_2} g_{jj},$ $i \in \{1,\ldots,n\}$

and deterministic protein production

$$\dot{x} = k_x \langle B \rangle \left(1 + \sum_{i=1}^{n_2} \sum_{j=1}^{i} g_{ij} \right). \tag{91}$$

Theorem 1 of [55] gives the time derivative of the expected value of any function $\varphi(x, s_{ij}, g_{ij})$ as

$$\begin{aligned} \frac{d\langle \varphi(x, s_{ij}, g_{ij}) \rangle}{dt} = & \left\langle \sum_{Events} \Delta\varphi(x, s_{ij}, g_{ij}) \times f(x, s_{ij}, g_{ij}) \right\rangle \\ & + \left\langle \frac{\partial \varphi(x, g_{ij})}{\partial x} k_x \langle B \rangle \left(1 + \sum_{i=1}^{n_2} \sum_{j=1}^{i} g_{ij} \right) \right\rangle, \end{aligned} \tag{92}$$

where the first term in the right hand side is contributed from stochastic events, and the second term is contributed from deterministic protein production. In this model, dynamics of $\langle x \rangle$, $\langle x s_{ij} \rangle$ and $\langle x g_{ij} \rangle$ are the same as equations (83) and (E.6), thus mean of protein, $\overline{\langle x s_{ij} \rangle}$, and $\overline{\langle x g_{ij} \rangle}$ will be equal to their value in previous section. Further, the second-order moment dynamics of protein can be added by selecting φ to be x^2 in equation (92)

$$\frac{d\langle x^2 \rangle}{dt} = 2k_x \langle B \rangle \left(\langle x \rangle + \left\langle \sum_{i=1}^{n_2} \sum_{j=1}^{i} x g_{ij} \right\rangle \right) - \frac{3k_2}{4} \left\langle \sum_{j=1}^{n_2} x^2 g_{jj} \right\rangle. \tag{93}$$

This equation is not closed thus we add dynamics of $\langle x^2 s_{ij} \rangle$, $i = \{1,\ldots,n_1\}$, $j = \{1,\ldots,i\}$ and

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$\langle x^2 g_{ij}\rangle,\ i=\{1,\dots,n_2\}\,,\ j=\{1,\dots,i\}$ to have a closed set of equations

$$\frac{d\langle x^2 s_{i1}\rangle}{dt} = 2k_x\langle B\rangle\langle x s_{i1}\rangle + \frac{k_2}{4}p_i\left\langle\sum_{j=1}^{n_2} x^2 g_{jj}\right\rangle - k_1\langle x^2 s_{i1}\rangle, \tag{94a}$$

$$\frac{d\langle x^2 s_{ij}\rangle}{dt} = 2k_x\langle B\rangle\langle x s_{ij}\rangle - k_1\langle x^2 s_{ij}\rangle + k_1\langle x^2 s_{(i-1)j}\rangle,\ j=\{2,\dots,i\}\,, \tag{94b}$$

$$\frac{d\langle x^2 g_{i1}\rangle}{dt} = 4k_x\langle B\rangle\langle x g_{i1}\rangle + k_1 p_i\left\langle\sum_{j=1}^{n_1} x^2 s_{jj}\right\rangle - k_2\langle x^2 g_{i1}\rangle, \tag{94c}$$

$$\frac{d\langle x^2 g_{ij}\rangle}{dt} = 4k_x\langle B\rangle\langle x g_{ij}\rangle - k_2\langle x^2 g_{ij}\rangle + k_2\langle x^2 g_{(i-1)j}\rangle, j=\{2,\dots,i\}\,. \tag{94d}$$

In order to calculate noise we need to express $\overline{\langle x^2 s_{ij}\rangle}$, and $\overline{\langle x^2 g_{ij}\rangle}$, which requires calculating the term $\overline{\langle\sum_{j=1}^{n_2} x^2 g_{jj}\rangle}$; this term can be derived by analyzing equation (93) in steady-state

$$\begin{aligned}&\frac{3k_2}{4}\overline{\left\langle\sum_{j=1}^{n_2} x^2 g_{jj}\right\rangle} = 2k_x\langle B\rangle\left(\overline{\langle x\rangle} + \overline{\left\langle\sum_{i=1}^{n_2}\sum_{j=1}^{i} x g_{ij}\right\rangle}\right) \Rightarrow\\ &\overline{\left\langle\sum_{j=1}^{n_2} x^2 g_{jj}\right\rangle} = \frac{4k_x^2\langle B\rangle^2\langle T_1\rangle\left((4-\beta)+\beta CV_{T_1}^2\right)}{3k_2} + \frac{16k_x^2\langle B\rangle^2\langle T_2\rangle\left((3-\beta)+(1-\beta)CV_{T_1}^2\right)}{3k_2},\end{aligned} \tag{95}$$

where in deriving this term we used equation (90) and we summed all the terms in equation (88). By having this term, we calculate $\overline{\langle x^2 s_{ij}\rangle}$ by recursion process. we derive $\overline{\langle x^2 s_{i1}\rangle}$ by substituting equation (95) in equation (94a). In the next step we use the definition of $\overline{\langle x^2 s_{i1}\rangle}$ to calculate $\overline{\langle x^2 s_{i2}\rangle}$ from equation (94b). We continue this process until we derive all the moments

$$\begin{aligned}\overline{\langle x^2 s_{ij}\rangle} =& \frac{k_x^2\langle B\rangle^2\langle T_1\rangle\left((4-\beta)+\beta CV_{T_1}^2\right)}{3k_1}p_i' + \frac{4k_x^2\langle B\rangle^2\langle T_2\rangle\left((3-\beta)+(1-\beta)CV_{T_2}^2\right)}{3k_1}p_i'\\ &+\frac{2k_x^2\langle B\rangle^2}{k_1}p_i'\left(\frac{\beta j^2+(2-\beta)j}{i}\right),\ i=\{1,\dots,n_1\},\ j=\{1,\dots,i\}.\end{aligned} \tag{96}$$

Expressing $\overline{\langle x^2 g_{ij}\rangle}$ requires calculation of the term $\overline{\langle\sum_{j=1}^{n_1} x^2 s_{jj}\rangle}$ which can be obtained from equation (96) as

$$\overline{\left\langle\sum_{j=1}^{n_1} x^2 s_{jj}\right\rangle} = \frac{4k_x^2\langle B\rangle^2\langle T_1\rangle\left((4-\beta)+\beta CV_{T_1}^2\right)}{3k_1} + \frac{4k_x^2\langle B\rangle^2\langle T_2\rangle\left((3-\beta)+(1-\beta)CV_{T_2}^2\right)}{3k_1}. \tag{97}$$

Thus $\overline{\langle x^2 g_{ij}\rangle}$ can be obtained with a recursion process from equations (94c) and (94d)

$$\begin{aligned}\overline{\langle x^2 g_{ij}\rangle} =& \frac{4k_x^2\langle B\rangle^2\langle\langle T_1\rangle\left((4-\beta)+\beta CV_{T_1}^2\right)}{3k_2}p_i + \frac{4k_x^2\langle B\rangle^2\langle T_2\rangle\left((3-\beta)+(1-\beta)CV_{T_2}^2\right)}{3k_2}p_i\\ &+\frac{8k_x^2\langle B\rangle^2}{k_2}p_i\left(\frac{(1-\beta)j^2+j}{i}\right),\ i=\{1,\dots,n_2\},\ j=\{1,\dots,i\}.\end{aligned} \tag{98}$$

Note that $\sum_{i=1}^{n_1}\sum_{j=1}^{i}\overline{\langle x^2 s_{ij}\rangle} + \sum_{i=1}^{n_2}\sum_{j=1}^{i}\overline{\langle x^2 g_{ij}\rangle} = \overline{\langle x^2\rangle}$ thus the second order moment of protein can be derived by adding all the terms in equations (96) and (98). $\overline{\langle x^2\rangle}$ can be simplified by using equations (7) and (18b) in the main article as

$$\overline{\langle x^2\rangle} = k_x^2\langle B\rangle^2\frac{4\langle T_1^3\rangle + 16\langle T_2^3\rangle + 2\langle T\rangle^3(3(2-\beta)^2+\beta^2(5-3\beta)CV_{T_1}^2 + 8(1-\beta)^2 CV_{T_2}^2)}{3\langle T\rangle}. \tag{99}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Finally, using the definition of CV^2 results in noise of protein raised from cell-cycle time variations

$$CV_E^2 = \frac{(4\langle T_1^3\rangle + 16\langle T_2^3\rangle)/\langle T\rangle^3 - 3(2-4\beta+\beta^2)^2}{3\left((\beta^2-4\beta+6)+\beta^2 CV_{T_1}^2 + 2(1-\beta)^2 CV_{T_2}^2\right)^2} \\ \frac{-3\beta^2(\beta^2(-2+CV_{T_1}^2))CV_{T_1}^2 - 4(\beta^2 CV_{T_1}^2 + (1-\beta)^2 CV_{T_2}^2)(2-12\beta+3\beta^2+3(\beta^2 CV_{T_1}^2 + (1-\beta)^2 CV_{T_2}^2))}{3\left((\beta^2-4\beta+6)+\beta^2 CV_{T_1}^2 + 2(1-\beta)^2 CV_{T_2}^2\right)^2}. \tag{100}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

E.3 Noise in protein count level contributed from random partitioning

In order to take into account noise caused by random partitioning of proteins between two daughter cells, we use the model shown in Figure S1C coupled with phase-type distributions. This model contains the following stochastic events

Event	Reset	Propensity
First phase-type evolution	$s_{ij}(t) \mapsto s_{ij}(t)-1,$ $s_{i(j+1)}(t) \mapsto s_{i(j+1)}(t)+1$	$k_1 s_{ij},$ $i \in \{2,\ldots,n_1\},$ $j \in \{1,\ldots,i-1\}$
Gene-duplication	$s_{jj}(t) \mapsto 0,$ $g_{i1}(t) \mapsto g_{i1}(t)+1$	$k_1 p'_i \sum_{j=1}^{n_1} s_{jj},$ $i \in \{1,\ldots,n_2\}$
Second phase-type evolution	$g_{ij}(t) \mapsto g_{ij}(t)-1,$ $g_{i(j+1)}(t) \mapsto g_{i(j+1)}(t)+1$	$k_2 g_{ij},$ $i \in \{2,\ldots,n_2\},$ $j \in \{1,\ldots,i-1\}$
Cell-division	$x(t_s) \mapsto x_+(t_s),$ $g_{jj}(t_s) \mapsto 0,$ $s_{i1}(t_s) \mapsto s_{i1}(t_s)+1$	$k_2 p_i \sum_{j=1}^{n_2} g_{jj},$ $i \in \{1,\ldots,n\}$

and deterministic protein production

$$\dot{x} = k_x \langle B \rangle \sum_{i=1}^{n_2} \sum_{j=1}^{i} g_{ij}. \tag{101}$$

Note that here x is a continuous random variable, hence x_+ is also obtained from a continious distribution. Connection between statistical statistical moments of x and x_+ is given by (5).

For this model, $\overline{\langle x \rangle}$, $\overline{\langle x s_{ij} \rangle}$, and $\overline{\langle x g_{ij} \rangle}$ are equal to their value in Section E.1 and Section E.2. However, dynamics of $\langle x^2 \rangle$ and $\langle x^2 s_{i1} \rangle$ are different

$$\frac{d\langle x^2 \rangle}{dt} = 2k_x \langle B \rangle \left(\langle x \rangle + \left\langle \sum_{i=1}^{n_2} \sum_{j=1}^{i} x g_{ij} \right\rangle \right) + \frac{1}{4} \alpha k_2 \left\langle \sum_{j=1}^{n_2} x g_{jj} \right\rangle - \frac{3k_2}{4} \left\langle \sum_{j=1}^{n_2} x^2 g_{jj} \right\rangle, \tag{102a}$$

$$\frac{d\langle x^2 s_{i1} \rangle}{dt} = 2k_x \langle B \rangle \langle x s_{i1} \rangle + \frac{k_2}{4} p'_i \left\langle \sum_{j=1}^{n_2} x^2 g_{jj} \right\rangle + \frac{1}{4} \alpha k_2 p'_i \left\langle \sum_{j=1}^{n_2} x g_{jj} \right\rangle - k_1 \langle x^2 s_{i1} \rangle, \tag{102b}$$

note that dynamics of $\langle x^2 s_{ij} \rangle$, $\langle x^2 g_{i1} \rangle$, and $\langle x^2 g_{ij} \rangle$ are identical to equations (94b), (94c), and (94d). Similar to previous section, we start by deriving the term $\overline{\langle \sum_{j=1}^{n} x^2 g_{jj} \rangle}$. Analyzing equation (102a) in steady-state gives this term as

$$\begin{aligned} \overline{\left\langle \sum_{j=1}^{n_2} x^2 g_{jj} \right\rangle} = & \frac{4k_x^2 \langle B \rangle^2 \langle T_1 \rangle \left((4-\beta) + \beta CV_{T_1}^2 \right)}{3k_2} \\ & + \frac{16k_x^2 \langle B \rangle^2 \langle T_2 \rangle \left((3-\beta) + (1-\beta) CV_{T_1}^2 \right)}{3k_2} + \frac{2\alpha k_x \langle B \rangle (2-\beta)}{3k_2}. \end{aligned} \tag{103}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Substituting equation (103) in equations (102b) and (94b) results in

$$\begin{aligned}\overline{\langle x^2 s_{ij}\rangle} =& \frac{k_x^2\langle B\rangle^2\langle T_1\rangle\left((4-\beta)+\beta CV_{T_1}^2\right)}{3k_1}p_i' + \frac{4k_x^2\langle B\rangle^2\langle T_2\rangle\left((3-\beta)+(1-\beta)CV_{T_2}^2\right)}{3k_1}p_i' \\ &+ \frac{2k_x^2\langle B\rangle^2}{k_1}p_i'\left(\frac{\beta j^2+(2-\beta)j}{i}\right) + \frac{2\alpha k_x(2-\beta)\langle B\rangle}{3k_1}p_i', \\ & i=\{1,\ldots,n_1\}\ j=\{1,\ldots,i\}.\end{aligned} \tag{104}$$

In the next step we derive moments $\overline{\langle x^2 g_{ij}\rangle}$; we start by calculating $\overline{\langle \sum_{j=1}^{n_1} x^2 s_{jj}\rangle}$ from (104)

$$\begin{aligned}\overline{\left\langle \sum_{j=1}^{n_1} x^2 s_{jj}\right\rangle} =& \frac{4k_x^2\langle B\rangle^2\langle T_1\rangle\left((4-\beta)+\beta CV_{T_1}^2\right)}{3k_1} \\ &+ \frac{4k_x^2\langle B\rangle^2\langle T_2\rangle\left((3-\beta)+(1-\beta)CV_{T_2}^2\right)}{3k_1} + \frac{2k_x\langle B\rangle(2-\beta)}{3k_1}.\end{aligned} \tag{105}$$

By having this term, the moments $\overline{\langle x^2 g_{ij}\rangle}$ are derived by solving equations (94c) and (94d) in steady-state

$$\begin{aligned}\overline{\langle x^2 g_{ij}\rangle} =& \frac{4k_x^2\langle B\rangle^2\langle\langle T_1\rangle\left((4-\beta)+\beta CV_{T_1}^2\right)}{3k_2}p_i + \frac{4k_x^2\langle B\rangle^2\langle T_2\rangle\left((3-\beta)+(1-\beta)CV_{T_2}^2\right)}{3k_2}p_i \\ &+ \frac{8k_x^2\langle B\rangle^2}{k_2}p_i\left(\frac{(1-\beta)j^2+j}{i}\right) + \frac{2\alpha k_x\langle B\rangle(2-\beta)}{3k_2}p_i, \\ & i=\{1,\ldots,n_2\},\ j=\{1,\ldots,i\}.\end{aligned} \tag{106}$$

Note that

$$\overline{\langle x^2\rangle} = \left\langle x^2\left(\sum_{i=1}^{n_1}\sum_{j=1}^{i} s_{ij} + \sum_{i=1}^{n_2}\sum_{j=1}^{i} g_{ij}\right)\right\rangle = \sum_{i=1}^{n_1}\sum_{j=1}^{i}\overline{\langle x^2 s_{ij}\rangle} + \sum_{i=1}^{n_2}\sum_{j=1}^{i}\overline{\langle x^2 g_{ij}\rangle}, \tag{107}$$

hence the second-order moment is

$$\begin{aligned}\overline{\langle x^2\rangle} =& \frac{4\langle T_1^3\rangle + 16\langle T_2^3\rangle + 2\langle T\rangle^3(3(2-\beta)^2+\beta^2(5-3\beta)CV_{T_1}^2 + 8(1-\beta)^2 CV_{T_2}^2)}{3\langle T\rangle} \\ &+ \frac{2\alpha k_x\langle B\rangle(2-\beta)\langle T\rangle}{3}.\end{aligned} \tag{108}$$

Coefficient of variation squared gives noise raised from partitioning and cell-cycle variations, which subtracting equation (100) from results gives partitioning noise as

$$CV_R^2 = \frac{4\alpha(2-\beta)}{3\left((\beta^2-4\beta+6)+\beta^2 CV_{T_1}^2 + 2(1-\beta)^2 CV_{T_2}^2\right)}\frac{1}{\overline{\langle x\rangle}}. \tag{109}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

E.4 Noise in protein count level contributed from stochastic production

In order to calculate the noise caused by stochastic birth of protein, we use the model introduced in Section C.1. For this model, moments dynamics of $\langle x^2\rangle$, $\langle x^2 s_{ij}\rangle$, and $\langle x^2 g_{ij}\rangle$ can be written as

$$\frac{d\langle x^2\rangle}{dt} = k_x\langle B^2\rangle(2-\beta) + 2k_x\langle B\rangle\left(\langle x\rangle + \left\langle \sum_{i=1}^{n_2}\sum_{j=1}^{i} x g_{ij}\right\rangle\right) + \frac{1}{4}\alpha k_2\left\langle\sum_{j=1}^{n_2} x g_{jj}\right\rangle - \frac{3k_2}{4}\left\langle\sum_{j=1}^{n_2} x^2 g_{jj}\right\rangle, \tag{110a}$$

$$\frac{d\langle x^2 s_{i1}\rangle}{dt} = \frac{k_x\langle B^2\rangle\beta p_i'}{i} + 2k_x\langle B\rangle\langle x s_{i1}\rangle + \frac{k_2}{4}p_i'\left\langle\sum_{j=1}^{n_2} x^2 g_{jj}\right\rangle + \frac{1}{4}\alpha k_2 p_i'\left\langle\sum_{j=1}^{n_2} x g_{jj}\right\rangle - k_1\langle x^2 s_{i1}\rangle, \tag{110b}$$

$$\frac{d\langle x^2 s_{ij}\rangle}{dt} = \frac{k_x\langle B^2\rangle\beta p_i'}{i} + 2k_x\langle B\rangle\langle x s_{ij}\rangle - k_1\langle x^2 s_{ij}\rangle + k_1\langle x^2 s_{(i-1)j}\rangle,\ j=\{2,\dots,i\}, \tag{110c}$$

$$\frac{d\langle x^2 g_{i1}\rangle}{dt} = \frac{2k_x\langle B^2\rangle(1-\beta)p_i}{i} + 4k_x\langle B\rangle\langle x g_{i1}\rangle + k_1 p_i\left\langle\sum_{j=1}^{n_1} x^2 s_{jj}\right\rangle - k_2\langle x^2 g_{i1}\rangle,, \tag{110d}$$

$$\frac{d\langle x^2 g_{ij}\rangle}{dt} = \frac{2k_x\langle B^2\rangle(1-\beta)p_i}{i} + 4k_x\langle B\rangle\langle x g_{ij}\rangle - k_2\langle x^2 g_{ij}\rangle + k_2\langle x^2 g_{(i-1)j}\rangle,\ j=\{2,\dots,i\}. \tag{110e}$$

The same as before we start by expressing the term $\overline{\langle\sum_{j=1}^{n_2} x^2 g_{jj}\rangle}$, this term is calculated by analyzing equation (110a) in steady-state

$$\begin{aligned}\overline{\left\langle\sum_{j=1}^{n_2} x^2 g_{jj}\right\rangle} =& \frac{4k_x^2\langle B\rangle^2\langle T_1\rangle\left((4-\beta)+\beta CV_{T_1}^2\right)}{3k_2} + \frac{2\alpha k_x\langle B\rangle(2-\beta)}{3k_2} \\ &+ \frac{16k_x^2\langle B\rangle^2\langle T_2\rangle\left((3-\beta)+(1-\beta)CV_{T_1}^2\right)}{3k_2} + \frac{4k_x(2-\beta)\langle B^2\rangle}{3k_2}.\end{aligned} \tag{111}$$

Substituting this term in equations (110b) and (110c) results in

$$\begin{aligned}\overline{\langle x^2 s_{ij}\rangle} =& \frac{k_x^2\langle B\rangle^2\langle T_1\rangle\left((4-\beta)+\beta CV_{T_1}^2\right)}{3k_1}p_i' + \frac{4k_x^2\langle B\rangle^2\langle T_2\rangle\left((3-\beta)+(1-\beta)CV_{T_2}^2\right)}{3k_1}p_i' \\ &+ \frac{2k_x^2\langle B\rangle^2}{k_1}p_i'\left(\frac{\beta j^2+(2-\beta)j}{i}\right) + \frac{2\alpha k_x(2-\beta)\langle B\rangle}{3k_1}p_i' \\ &+ \frac{k_x\langle B^2\rangle}{k_1}\left(\frac{2-\beta}{3}+\beta\frac{j}{i}\right)p_i',\ i=\{1,\dots,n_1\},\ j=\{1,\dots,i\}.\end{aligned} \tag{112}$$

Similar to previous section, solving equations (110d) and (110e) gives the $\overline{\langle x^2 g_{ij}\rangle}$

$$\begin{aligned}\overline{\langle x^2 g_{ij}\rangle} =& \frac{4k_x^2\langle B\rangle^2\langle\langle T_1\rangle\left((4-\beta)+\beta CV_{T_1}^2\right)}{3k_2}p_i + \frac{4k_x^2\langle B\rangle^2\langle T_2\rangle\left((3-\beta)+(1-\beta)CV_{T_2}^2\right)}{3k_2}p_i \\ &+ \frac{8k_x^2\langle B\rangle^2}{k_2}p_i\left(\frac{(1-\beta)j^2+j}{i}\right) + \frac{2\alpha k_x\langle B\rangle(2-\beta)}{3k_2}p_i \\ &+ \frac{2k_x\langle B^2\rangle}{k_2}\left(\frac{1+\beta}{3}+(1-\beta)\frac{j}{i}\right)p_i,\ i=\{1,\dots,n_2\},\ j=\{1,\dots,i\}.\end{aligned} \tag{113}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Finally summing all the moments $\overline{\langle x^2 s_{ij}\rangle}$, and $\overline{\langle x^2 g_{ij}\rangle}$ results in $\overline{\langle x^2\rangle}$ as

$$\begin{aligned}\overline{\langle x^2\rangle} =& \frac{4\langle T_1^3\rangle + 16\langle T_2^3\rangle + 2\langle T\rangle^3(3a(-(\beta+1)\beta CV_{T_1}^2 + a - 4) + 8CV_T^2 + 12)}{3\langle T\rangle} \\ &+ \frac{2\alpha k_x\langle B\rangle(2-a)\langle T\rangle}{3} + k_x\langle B^2\rangle\left(\frac{2-\beta}{3} + \beta\left(\frac{1+CV_{T_1}^2}{2}\right)\right)\langle T_1\rangle \\ &+ 2k_x\langle B^2\rangle\left(\frac{1+\beta}{3} + (1-\beta)\left(\frac{1+CV_{T_2}^2}{2}\right)\right)\langle T_2\rangle.\end{aligned} \tag{114}$$

Steady-state analysis gives the noise from stochastic birth, random partitioning, and cell-cycle time variations. Subtracting noise of cell-cycle time and partitioning in equations (100) and (109) results in noise caused by stochastic production of protein

$$CV_P^2 = \frac{(10 - 8\beta + 3\beta^2) + 6(1-\beta)^2 CV_{T_2}^2 + 3\beta^2 CV_{T_1}^2}{3\left((\beta^2 - 4\beta + 6) + \beta^2 CV_{T_1}^2 + 2(1-\beta)^2 CV_{T_2}^2\right)}\frac{\langle B^2\rangle}{\langle B\rangle}\frac{1}{\overline{\langle x\rangle}}. \tag{115}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

E.5 Effect of gene-duplication time in intrinsic noise

We investigate how the noise contributions from random partitioning and stochastic expression (CV_R^2 and CV_P^2 terms in equation (34) of the main text) change as β is varied between 0 and 1. Results show that CV_R^2 and CV_P^2 follow the same qualitative shapes as reported in Fig. 6. There exists a β^*

$$\beta^* = \frac{-\sqrt{2(2CV_{T_1}^4 + 5CV_{T_1}^2 CV_{T_2}^2 + 3CV_{T_1}^2 + 2CV_{T_2}^4 + 3CV_{T_2}^2 + 1)} + 2CV_{T_1}^2 + 4CV_{T_2}^2 + 2}{CV_{T_1}^2 + 2CV_{T_2}^2 + 1}, \tag{116}$$

such that CV_P^2 is minimized and CV_R^2 is maximized when $\beta = \beta^*$. Note that when $CV_{T_1}^2 = CV_{T_2}^2 = 0$, $\beta^* = 2 - \sqrt{2}$ as reported in the main text. The minimum value of CV_P^2 and the maximum value of CV_R^2 are given by

$$CV_P^2 = \frac{CV_{T_1}^2(3CV_{T_2}^2 + 7) - \sqrt{2(2CV_{T_1}^2 + CV_{T_2}^2 + 1)(CV_{T_1}^2 + 2CV_{T_2}^2 + 1)} + 7CV_{T_2}^2 + 3}{3(CV_{T_1}^2(CV_{T_2}^2 + 3) + 3CV_{T_2}^2 + 1)} \frac{\langle B^2 \rangle}{\langle B \rangle} \frac{1}{\overline{\langle x \rangle}}, \tag{117}$$

$$CV_R^2 = \frac{\sqrt{2}\alpha}{3\sqrt{(2CV_{T_1}^2 + CV_{T_2}^2 + 1)(CV_{T_1}^2 + 2CV_{T_2}^2 + 1)} - 3\sqrt{2}CV_{T_1}^2 - 3\sqrt{2}CV_{T_2}^2}, \tag{118}$$

respectively. Plots of β^* and optimal value of CV_R^2 and CV_P^2 as a function of $CV_{T_1}^2$ are shown in Fig. S4. Note that if noise in T_1 is high and T_2 is deterministic then β^* shifts towards zero. Similarly, if noise in T_2 is high and T_1 is deterministic then β^* shifts towards one.

F Noise level in unstable protein

Consider an unstable protein with sufficiently high degradation rate γ_x such that the protein level reaches steady-state instantaneously compared to the cell-cycle time (Fig. S4). Let τ denote the time from the last division event, then

$$\begin{aligned} \overline{\langle x|\tau < T_1 \rangle} &= \frac{k_x \langle B \rangle}{\gamma_x}, \\ \overline{\langle x|\tau > T_1 \rangle} &= \frac{2k_x \langle B \rangle}{\gamma_x}, \end{aligned} \tag{119}$$

where T_1 is the time in which duplication happens. The mean level of an unstable protein can be calculated as

$$\overline{\langle x \rangle} = \overline{\langle x|\tau < T_1 \rangle} p(\tau < T_1) + \overline{\langle x|\tau > T_1 \rangle} p(\tau > T_1), \tag{120}$$

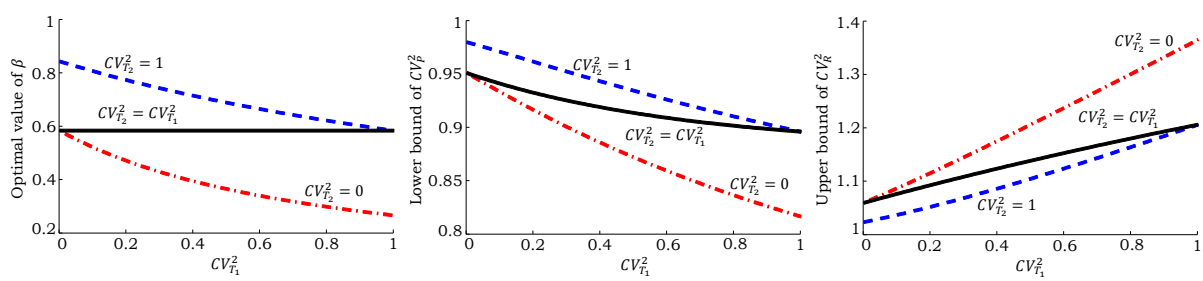

Figure S3: **Effect of gene-duplication on intrinsic noise level**. *Left*: Value of β where CV_P^2 is minimized and CV_R^2 is maximized as a function of $CV_{T_1}^2$. When $CV_{T_1}^2 = CV_{T_2}^2$, noise levels always reach their extrema at $\beta = 2 - \sqrt{2}$. *Middle & Right*: Extremum values of CV_P^2 and CV_R^2 as a functions of $CV_{T_1}^2$. Noise levels are normalized by their values at $\beta = 0$.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where $p(\tau < T_1)$ and $p(\tau > T_1)$ denote the probability of being in the time interval before and after gene-duplication. Using

$$p(\tau < T_1) = \beta, \quad p(\tau > T_1) = (1-\beta), \tag{121}$$

we obtain

$$\overline{\langle x \rangle} = \frac{k_x \langle B \rangle (2-\beta)}{\gamma_x}. \tag{122}$$

To compute the extrinsic noise component we consider deterministic protein production and decay. The second-order moment of $x(t)$ is given by

$$\begin{aligned} \overline{\langle x^2 | \tau < T_1 \rangle} &= \left(\frac{k_x \langle B \rangle}{\gamma_x}\right)^2 \\ \overline{\langle x^2 | \tau > T_1 \rangle} &= \left(\frac{2k_x \langle B \rangle}{\gamma_x}\right)^2 \end{aligned} \Rightarrow \overline{\langle x^2 \rangle} = \left(\frac{k_x \langle B \rangle}{\gamma_x}\right)^2 \beta + \left(\frac{2k_x \langle B \rangle}{\gamma_x}\right)^2 (1-\beta). \tag{123}$$

By using definition of CV^2, extrinsic noise is

$$CV_E^2 = \frac{(1-\beta)\beta}{(2-\beta)^2}, \tag{124}$$

which is zero at $\beta = 0, 1$ and reaches its maximum at $\beta = 2/3$ (Fig. S4).

Next we compute the intrinsic noise component. If the protein decay is sufficiently high, the noise contribution from partitioning errors will be negligible because any errors will be instantaneously corrected due to rapid protein turnover. Noise raised from stochastic gene expression can be investigated by considering a model containing stochastic bursty production and stochastic degradation of proteins, where after gene-duplication the burst frequency doubles. Again assuming large enough γ_x, $\overline{\langle x^2 | \tau < T_1 \rangle}$ is equal to the steady-state second-order moment of a stochastic model with burst frequency k_x (analyzed in [64])

$$\overline{\langle x^2 | \tau < T_1 \rangle} = \left(\frac{k_x \langle B \rangle}{\gamma_x}\right)^2 + \frac{k_x \langle B^2 \rangle}{2\gamma_x^2} + \frac{k_x \langle B \rangle}{2\gamma_x}. \tag{125}$$

In comparison with equation (123), there are two extra terms at the right hand side of $\overline{\langle x^2 | \tau < T_1 \rangle}$. The first extra term is due to production of protein in random bursts and the second one is due to stochastic degradation of protein molecules. Further for the same reasons (large degradation rate and rapid equilibration of the distribution), $\overline{\langle x^2 | \tau > T_1 \rangle}$ is equal to the second-order moment of a model containing stochastic bursty production of proteins with burst frequency $2k_x$ which is

$$\overline{\langle x^2 | \tau > T_1 \rangle} = \left(\frac{2k_x \langle B \rangle}{\gamma_x}\right)^2 + \frac{k_x \langle B^2 \rangle}{\gamma_x^2} + \frac{k_x \langle B \rangle}{\gamma_x}. \tag{126}$$

Thus the second order moment of an unstable protein can be written as

$$\begin{aligned} \overline{\langle x^2 \rangle} = &\left(\frac{k_x \langle B \rangle}{\gamma_x}\right)^2 \beta + \frac{k_x \langle B^2 \rangle}{2\gamma_x^2}\beta + \frac{k_x \langle B \rangle}{2\gamma_x}\beta \\ &+ \left(\frac{2k_x \langle B \rangle}{\gamma_x}\right)^2 (1-\beta) + \frac{k_x \langle B^2 \rangle}{\gamma_x^2}(1-\beta) + \frac{k_x \langle B \rangle}{\gamma_x}(1-\beta). \end{aligned} \tag{127}$$

Using definition of CV^2 and subtracting extrinsic noise we obtain the following noise contribution from stochastic expression and decay

$$CV_P^2 = \frac{1}{2}\left(\frac{\langle B^2 \rangle}{\langle B \rangle} + 1\right)\frac{1}{\overline{\langle x \rangle}}. \tag{128}$$

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

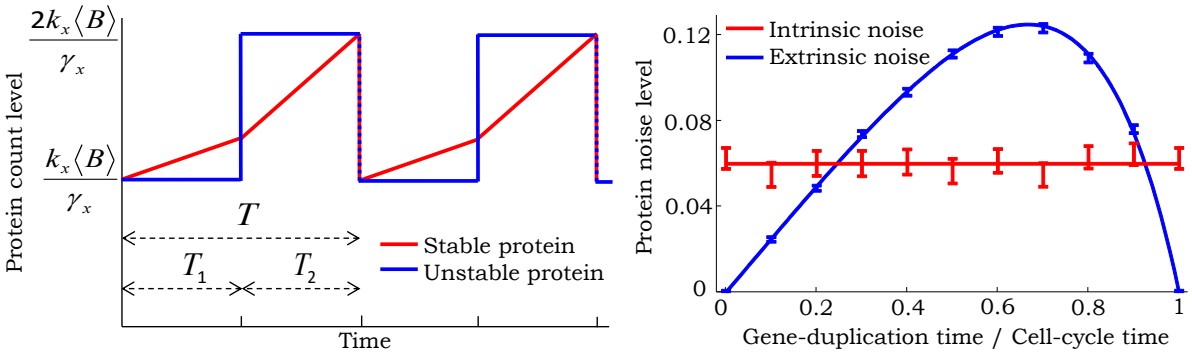

Figure S4: **Contribution of gene duplication to noise levels of an unstable protein.** *left*: For a stable protein, copy numbers accumulate in a bilinear fashion. In contrast, an unstable protein reaches equilibrium rapidly and its level changes in steps. *Right*: Extrinsic and intrinsic noise predicted for an unstable protein as a function of β. Solid lines are predictions from (124) and (128), which agree with estimates from $20,000$ Monte Carlo simulations. Parameters taken as $\gamma_x = 10hr^{-1}$, and a geometric burst with $\langle B\rangle = 6$. Burst frequency is changed to have a constant mean protein level of 100 molecules for different values of β. 95% confidence intervals are calculated via bootstrapping.

Acknowledgements

AS is supported by the National Science Foundation Grant DMS-1312926.

References

1. Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. Nature 422: 633-637.

2. Raser JM, O'Shea EK (2005) Noise in gene expression: Origins, consequences, and control. Science 309: 2010-2013.

3. Neuert G, Munsky B, Tan RZ, Teytelman L, Khammash M, et al. (2013) Systematic identification of signal-activated stochastic gene regulation. Science 339: 584-587.

4. Libby E, Perkins TJ, Swain PS (2007) Noisy information processing through transcriptional regulation. Proceedings of the National Academy of Sciences 104: 7151-7156.

5. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB (2004) Noise minimization in eukaryotic gene expression. PLOS Biology 2: e137.

6. Lehner B (2008) Selection to minimise noise in living systems and its implications for the evolution of gene expression. Molecular Systems Biology 4: 170.

7. Losick R, Desplan C (2008) Stochasticity and cell fate. Science 320: 65-68.

8. Arkin A, Ross J, McAdams HH (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. Genetics 149: 1633–1648.

9. Weinberger L, Burnett J, Toettcher J, Arkin A, Schaffer D (2005) Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. Cell 122: 169-182.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10. Weinberger LS, Dar RD, Simpson ML (2008) Transient-mediated fate determination in a transcriptional circuit of HIV. Nature Genetics 40: 466-470.

11. Singh A, Weinberger LS (2009) Stochastic gene expression as a molecular switch for viral latency. Current Opinion in Microbiology 12: 460-466.

12. Dar RD, Hosmane NN, Arkin MR, Siliciano RF, Weinberger LS (2014) Sreening for noise in gene expression identifies drug synergies. Science 344: 1392-1396.

13. Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. Nature 467: 167-173.

14. Veening JW, Smits WK, Kuipers OP (2008) Bistability, epigenetics, and bet-hedging in bacteria. Annual Review of Microbiology 62: 193210.

15. Kussell E, Leibler S (2005) Phenotypic diversity, population growth, and information in fluctuating environments. Science 309: 2075-2078.

16. Balaban N, Merrin J, Chait R, Kowalik L, Leibler S (2004) Bacterial persistence as a phenotypic switch. Science 305: 1622-1625.

17. Snchez-Romero MA, Casadess J (2014) Contribution of phenotypic heterogeneity to adaptive antibiotic resistance. Proceedings of the National Academy of Sciences 111: 355–360.

18. Neildez-Nguyen TMA, Parisot A, Vignal C, Rameau P, Stockholm D, et al. (2008) Epigenetic gene expression noise and phenotypic diversification of clonal cell populations. Differentiation 76: 33–40.

19. Paldi A (2003) Stochastic gene expression during cell differentiation: order from disorder? Cellular and Molecular Life Sciences 60: 1775-1778.

20. Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135: 216–226.

21. Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: from theories to phenotypes. Nature Reviews Genetics 6: 451-464.

22. Magklara A, Lomvardas S (2013) Stochastic gene expression in mammals: lessons from olfaction. Trends in Cell Biology 23: 449–456.

23. Munsky B, B Trinh B, Khammash M (2009) Listening to the noise: random fluctuations reveal gene network parameters. Molecular systems biology 5: 318.

24. Wang P, Robert L, Pelletier J, Dang WL, Taddei F, et al. (2010) Robust growth of *Escherichia coli*. Current biology 20: 1099–1103.

25. Lambert G, Kussell E (2015) Quantifying selective pressures driving bacterial evolution using lineage analysis. Physical Review X 5: 011016.

26. Tsukanov R, Reshes G, Carmon G, Fischer-Friedrich E, Gov NS, et al. (2011) Timing of Z-ring localization in *Escherichia coli*. Physical Biology 8: 066003.

27. Reshes G, Vanounou S, Fishov I, Feingold M (2008) Cell shape dynamics in *Escherichia coli*. Biophysical Journal 94: 251–264.

28. Reshes G, Vanounou S, Fishov I, Feingold M (2008) Timing the start of division in *E. coli*: a single-cell study. Physical Biology 5: 046001.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

29. Roeder A, Chickarmane V, Obara B, Manjunath B, Meyerowitz EM (2010) Variability in the control of cell division underlies sepal epidermal patterning in Arabidopsis thaliana. PLOS Biology 8: e1000367.

30. Zilman A, Ganusov V, Perelson A (2010) Stochastic models of lymphocyte proliferation and death. PLOS ONE 5: e12775.

31. Hawkins ED, Markham JF, McGuinness LP, Hodgkin P (2009) A single-cell pedigree analysis of alternative stochastic lymphocyte fates. Proceedings of the National Academy of Sciences 106: 13457-13462.

32. Stukalin EB, Aifuwa I, Kim JS, Wirtz D, Sun SX (2013) Age-dependent stochastic models for understanding population fluctuations in continuously cultured cells. Journal of The Royal Society Interface 10.

33. Huh D, Paulsson J (2011) Random partitioning of molecules at cell division. Proceedings of the National Academy of Sciences 108: 15004–15009.

34. Gonze D (2013) Modeling the effect of cell division on genetic oscillators. Journal of Theoretical Biology 325: 22–33.

35. Lloyd-Price J, Tran H, Ribeiro AS (2014) Dynamics of small genetic circuits subject to stochastic partitioning in cell division. Journal of Theoretical Biology 356: 11-19.

36. Zopf CJ, Quinn K, Zeidman J, Maheshri N (2013) Cell-cycle dependence of transcription dominates noise in gene expression. PLOS Computational Biology 9: e1003161.

37. Narula J, Kuchina A, Lee DyD, Fujita M, Sel GM, et al. (2015) Chromosomal arrangement of phosphorelay genes couples sporulation and DNA replication. Cell 162: 328–337.

38. Schwabe A, Bruggeman FJ (2014) Contributions of cell growth and biochemical reactions to non-genetic variability of cells. Biophysical Journal 107: 301–313.

39. Huh D, Paulsson J (2011) Non-genetic heterogeneity from stochastic partitioning at cell division. Nature Genetics 43: 95–100.

40. Antunes D, Singh A (2014) Quantifying gene expression variability arising from randomness in cell division times. Journal of Mathematical Biology 71: 1–27.

41. Yu J, Xiao J, Ren X, Lao K, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. Science 311: 1600-1603.

42. Paulsson J (2005) Model of stochastic gene expression. Physics of Life Reviews 2: 157-175.

43. Shahrezaei V, Swain PS (2008) Analytical distributions for stochastic gene expression. Proceedings of the National Academy of Sciences 105: 17256-17261.

44. Singh A, Hespanha JP (2009) Optimal feedback strength for noise suppression in autoregulatory gene networks. Biophysical Journal 96: 4013–4023.

45. Jia T, Kulkarni RV (2011) Intrinsic noise in stochastic models of gene expression with molecular memory and bursting. Journal of Mathematical Biology 106: 058102.

46. Alon U (2006) An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman and Hall/CRC.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

47. Taniguchi Y, Choi P, Li G, Chen H, Babu M, et al. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 329: 533-538.

48. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, et al. (2011) Global quantification of mammalian gene expression control. Nature 473: 337–342.

49. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. Proceedings of the National Academy of Sciences 99: 12795-12800.

50. Berg OG (1978) A model for the statistical fluctuations of protein numbers in a microbial population. Journal of Theoretical Biology 71: 587–603.

51. Rigney DR (1979) Stochastic model of constitutive protein levels in growing and dividing bacterial cells. Journal of Theoretical Biology 76: 453–480.

52. Singh A, Hespanha JP (2010) Stochastic hybrid systems for studying biochemical processes. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 368: 4995-5011.

53. Daigle BJ, Soltani M, Petzold LR, Singh A (2015) Inferring single-cell gene expression mechanisms using stochastic simulation. Bioinformatics 31: 1428–1435.

54. Tijms HC (1994) Stochastic models: an algorithmic approach. John Wiley & Sons.

55. Hespanha JP, Singh A (2005) Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems. International Journal of Robust and Nonlinear Control 15: 669–689.

56. Singh A, Hespanha JP (2011) Approximate moment dynamics for chemically reacting systems. IEEE Transactions on Automatic Control 56: 414-418.

57. Gomez-Uribe CA, Verghese GC (2007) Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. Journal of Chemical Physics 126: 024109.

58. Lee CH, Kim K, Kim P (2009) A moment closure method for stochastic reaction networks. Journal of Chemical Physics 130: 134107.

59. Goutsias J (2007) Classical versus stochastic kinetics modeling of biochemical reaction systems. Biophysical Journal 92: 2350-2365.

60. Gillespie CS (2009) Moment-closure approximations for mass-action models. IET systems biology 3: 52–58.

61. Soltani M, Vargas-Garcia CA, Singh A (2015) Conditional moment closure schemes for studying stochastic dynamics of genetic circuits. IEEE Transactions on Biomedical Circuits and Systems : Accepted to publish.

62. Wang H, Yuan Z, Liu P, Zhou T (2015) Division time-based amplifiers for stochastic gene expression. Molecular BioSystems 11: 2417-2428.

63. Hilfinger A, Paulsson J (2011) Separating intrinsic from extrinsic fluctuations in dynamic biological systems. Proceedings of the National Academy of Sciences 108: 12167-12172.

64. Singh A, Soltani M (2013) Quantifying intrinsic and extrinsic variability in stochastic gene expression models. PLOS ONE 8: e84301.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

65. Shahrezaei V, Ollivier JF, Swain PS (2008) Colored extrinsic fluctuations and stochastic gene expression. Molecular Systems Biology 4.

66. Scott M, Ingalls B, Kaern M (2006) Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. Chaos 16: 026107.

67. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nature Genetics 31: 69–73.

68. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, et al. (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature Genetics 441: 840-846.

69. Singh A, Razooky B, Cox CD, Simpson ML, Weinberger LS (2010) Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. Biophysical Journal 98: L32-L34.

70. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, et al. (2006) Noise in protein expression scales with natural protein abundance. Nature Genetics 38: 636-643.

71. Friedman N, Cai L, Xie X (2006) Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Physical Review Letters 97: 168302.

72. Paulsson J (2005) Model of stochastic gene expression. Physics of Life Reviews 2: 157–175.

73. Sherman MS, Cohen BA (2014) A computational framework for analyzing stochasticity in gene expression. PLOS Computational Biology 10: e1003596.

74. Golding I, Paulsson J, Zawilski S, Cox E (2005) Real-time kinetics of gene activity in individual bacteria. Cell 123: 1025-1036.

75. Yu J, Xiao J, Ren X, Lao K, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. Science 311: 1600-1603.

76. Cai L, Xie NFXS (2006) Stochastic protein expression in individual cells at the single molecule level. Nature 440: 358-362.

77. Kumar N, Singh A, Kulkarni RV (2015) Transcriptional bursting in gene expression: analytical results for general stochastic models. To appear in PLOS Computational Biology .

78. Singh A (2014) Transient changes in intercellular protein variability identify sources of noise in gene expression. Biophysical Journal 107: 2214–2220.

79. Osella M, Nugent E, Lagomarsino MC (2014) Concerted control of *Escherichia coli* cell division. Proceedings of the National Academy of Sciences 111: 3431-3435.

80. Robert L, Hoffmann M, Krell N, Aymerich S, Robert J, et al. (2014) Division in *Escherichia coli* is triggered by a size-sensing rather than a timing mechanism. BMC Biology 12: 17.

81. Kafri R, Levy J, Ginzberg MB, Oh S, Lahav G, et al. (2013) Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. Nature 494: 480–483.

82. Amir A (2014) Cell size regulation in bacteria. Physical Review Letters 112: 208102.

83. Taheri-Araghi S, Bradde S, Sauls JT, Hill NS, Levin PA, et al. (2015) Cell-size control and homeostasis in bacteria. Current Biology 25: 385-391.

bioRxiv preprint doi: https://doi.org/10.1101/026559; this version posted October 3, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

84. Campos M, Surovtsev IV, Kato S, Paintdakhi A, Beltran B, et al. (2014) A constant size extension drives bacterial cell size homeostasis. Cell 159: 1433-1446.

85. Padovan-Merhar O, Nair GP, Biaesch AG, Mayer A, Scarfone S, et al. (2015) Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptiona mechanisms. Molecular Cell 58: 339–352.

86. Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, et al. (2011) Mammalian genes are transcribed with widely different bursting kinetics. Science 332: 472-474.

87. Brown CR, Mao C, Falkovskaia E, Jurica MS, Boeger H (2013) Linking stochastic fluctuations in chromatin structure and gene expression. PLOS Biology 11: e1001621.

88. Raj A, Peskin C, Tranchina D, Vargas D, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. PLOS Biology 4: e309.

89. Hornung G, Bar-Ziv R, Rosin D, Tokuriki N, Tawfik DS, et al. (2012) Noise-mean relationship in mutated promoters. Genome Research 22: 2409-2417.

90. Singh A, Razooky BS, Dar RD, Weinberger LS (2012) Dynamics of protein noise can distinguish between alternate sources of gene-expression variability. Molecular Systems Biology 8: 607.

91. Dar RD, Razooky BS, Singh A, Trimeloni T, McCollum J, et al. (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. Proceedings of the National Academy of Sciences 109: 17454–17459.

92. Corrigan AM, Chubbemail JR (2014) Regulation of transcriptional bursting by a naturally oscillating signal. Current Biology 24: 205-211.

93. Bothma JP, Garcia HG, Esposito E, Schlissel G, Gregor T, et al. (2014) Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living drosophila embryos. Proceedings of the National Academy of Sciences 111: 10598–10603.

94. Chubb JR, Trcek T, Shenoy SM, Singer RH (2006) Transcriptional pulsing of a developmental gene. Current Biology 16: 1018–1025.

95. Chong S, Chen C, Ge H, Xie XS (2014) Mechanism of transcriptional bursting in bacteria. Cell 158: 314-326.

96. McQuarrie DA (1967) Stochastic approach to chemical kinetics. Journal of Applied Probability 4: 413–478.

97. Gillespie DT (2001) Approximate accelerated stochastic simulation of chemically reacting systems. Journal of Chemical Physics 115: 1716–1733.